Author:
Hales, Jack

Title:
Investigation and comparison of the human interactions of flaviviral NS5 protein

Investigation and comparison of the human interactions of flaviviral NS5 protein

JACK HALES

SCHOOL: CELLULAR AND MOLECULAR MEDICINE

SUPERVISOR: DR ANDREW DAVIDSON

START DATE: 22/09/2016

This project is my own work except where indicated. All text, figures, tables, data, or results which are not my own work, are indicated and the sources acknowledged. In addition, I confirm that the hardcopy and the e-submission are identical.

Signed:_____ Dated:_____

In memory of:

Richard John Hales

(1945 – 2017)

Who part funded this master's degree

And instilled the values of hard work and doing what you love

Acknowledgements

I would like to acknowledge first the input of Dr Andrew Davidson who supervised this project. Having been my supervisor through both my undergraduate project and master's project, there is no doubt that his patient guidance and support throughout the years have been instrumental in developing my skills and helping me to progress onto a PhD project and into future research. For that, I am extremely grateful.

I would also like to acknowledge Dr Lisa Stevens and Mr Joshua Lee for their input at the beginning of this project (as stated within) and also their advice and support during my time in the lab.

Finally, I would like to acknowledge the constant support from family throughout the duration of this project and beyond.

Table of contents

Table of figures and tables

Abbreviations

aa – Amino acid

Ab – Antibody

ADE – Antibody-dependent enhancement

ADEM – Acute disseminated encephalomyelitis

ALKV – Alkhurma virus

ALT – Alanine aminotransferase

AST – Aspartate aminotransferase

AXL receptor – Tyrosine-protein kinase receptor UFO

B-MERC – β-Mercaptoethanol

BSA – Bovine serum albumin

C protein – Capsid protein

CSF – Cerebrospinal fluid

CyPA – Cyclophilin A

C3/4/5 – ZNS5-BirA clone 3/4/5

D2NS5 – Dengue virus serotype 2 NS5 protein

D4NS5 – Dengue virus serotype 4 NS5 protein

DALS – Disability adjusted life scale

DAVID – Database for annotation, visualisation, and integrated discovery

DC – Dendritic cell

DC-SIGN - Dendritic cell-specific intracellular adhesion molecule-3-grabbing non-integrin

DENV – Dengue virus

DF – Dengue fever

DHF – Dengue haemorrhagic fever

DSS – Dengue shock syndrome

E protein – Envelope protein

ECM – Extracellular matrix

EF-1α – Eukaryotic translation elongation factor 1α

ER – Endoplasmic reticulum

FBS – Foetal bovine serum

FDR – False discovery rate

FRT – Flp recombination target

GBS – Guillain-Barré syndrome

GM – Growth media

GO – Gene ontology

GPI - Glycosyl-phosphatidylinositol

HEK – Human embryonic kidney

IFA – Immunofluorescence assay

IFIT1 – Interferon induced protein with tetratricopeptide repeats 1

IFN – Interferon

IFNAR1/2 – Interferon alpha and beta receptor subunit 1/2

iNOS – Inducible nitric oxide synthase

IRES – Internal ribosome entry site

IRF3 – Interferon regulatory transcription factor 3

ISRE – Interferon-stimulated response elements

JEV – Japanese encephalitis virus

KO – Knock-out

LPS - Lipopolysaccharide

M protein – Membrane protein

MAC – Membrane attack complex

MAPK – Mitogen-activated protein kinase

MAVS – Mitochondrial antiviral-signalling protein

MDCK – Modoc virus

MERS-CoV – Middle East respiratory coronavirus

MS – Mass spectrometry

MTase – Methyltransferase

MuRep - Dengue replicon-containing/expressing cells with persistent mutation introduced

NS3 – Non-structural protein 3

NS5 – Non-structural protein 5

NTPase – Nucleotide triphosphatase

ORF – Open reading frame

PAFR – Platelet-activating factor receptor

PBS – Phosphate-buffered saline

PD – Proteome Discoverer

prM protein – Pre-membrane protein

PRMT5 – Protein arginine methyltransferase 5

PTB – Polypyrimidine tract-binding protein

RdRp – RNA-dependent RNA polymerase

Rep – Dengue replicon-containing/expressing cells

RIG-I – Retinoic acid-inducible gene I

ROS – Reactive oxygen species

SILAC – Stable isotope labelling with amino acids in cell culture

SLA – Stem loop A

STAT1/2 – Signal transducer and activator of transcription ½

STING – Stimulator of interferon genes

STRING – Search Tool for the Retrieval of Interacting Genes/Proteins

TAM receptors – Tyro3, Axl, and Mer receptors

TBK1 – Tank binding kinase 1

TBEV – Tick-borne encephalitis virus

TIAR – TIA1 cytotoxic granule associated RNA binding protein like 1

TIA1 – TIA1 cytotoxic granule associated RNA binding protein

TIM receptor – T-cell immunoglobulin mucin domain receptor

TLR – Toll-like receptor

TMT – Tandem mass tag

Tyro3 receptor – Tyrosine-protein kinase receptor TYRO3

UFP – Unfolded protein

UTR – Untranslated region

WT – Wild-type

ZF – Zika fever

ZIKV – Zika virus

ZNS5 – Zika virus non-structural protein 5

293-Flp – HEK 293-Flp cells

Abstract

Dengue virus and Zika virus are both emerging flaviviruses and pose growing risks to the global human population. Despite numerous links being drawn on the similarities of both viruses, there have been no direct, large-scale comparisons of the interactomes of these two viruses. With NS5 playing such an important central role with virus-host interactions, the similarities between this protein's interactors could advance our understanding of the evolutionary closeness between the species and inform future studies about the potential for dengue-developed therapeutic host targets to be utilised for Zika fever. It would also provide a side-by-side comparison to show the differences in interactome that may explain the biological differences between dengue and Zika virus. Using transfected HEK-293-Flp cells as a viral gene expression system and pulldown LC-MS/MS proteomics, the similarities and differences between ZNS5 and DENV-NS5 interactors are explored, revealing that differences lie largely at the individual protein level with mostly similar pathways being targeted by both viruses. The comparability of different proteomic programmes and pulldown techniques was also investigated with a comparison between MaxQuant and Proteome Discoverer revealing large differences in the data generated from the sample spectra at both the protein identification level and in large scale pathways analysis with STRING and DAVID. BioID-labelling and FLAG-tag pulldowns were also compared with the intention to investigate the differences in interactomes generated by these two methods which once again highlighted the potential for differences in LC-MS/MS proteomics to arise. Cdk1 and UBR5, two ZNS5-specific interactors detected robustly by LC-MS/MS proteomics, were then validated as interactors in reverse-pulldowns and western blots to test the reliability of LC-MS/MS proteomics but called into question the quantitative aspects of the technique by also pulling down DENV-NS5 which further highlighted current limitations in predictive proteomics.

1: Introduction

1.1: Flaviviruses:

The *Flavivirus* genus is a taxonomic collection of >50 arthropod-borne viruses that lies within the family *Flaviviridae* alongside *Pestiviruses, Pegiviruses,* and *Hepaciviruses*. Of the known flaviviruses, roughly half of them cause significant clinical disease in humans (1) including diseases of high global importance such as dengue fever (DF), West Nile fever, yellow fever, Japanese encephalitis, tick-borne encephalitis, and Zika fever (ZF). However flavivirus infection is not exclusively limited to human infection with other flaviviruses capable of causing disease in cattle, pigs, birds, primates, and rodents. Most viruses within the *Flavivirus* genus are transmitted *via* an arthropod vector, generally mosquitoes or ticks, although there are some viruses where no current vector transmission is known between animals (direct transmission of the virus) and some that solely transmit through the arthropod population. There are also those who can transmit *via* multiple routes like Zika virus with mosquitoes acting as the major route for transmission but with the additional capability to be sexually transmitted (2). All flaviviruses share the characteristics of being morphologically-similar, enveloped viruses that contain a single-stranded, positive sense RNA genome which contains the same general gene organisation across all viruses (3). While it was previously believed that the bulk of the mosquito-borne flaviviruses could be further categorised into two main groups; the haemorrhagic, *Aedes*-borne viruses and the neurologic, *Culex*-borne viruses (4), the recent emergence of neurological side effects from infection with the *Aedes*-borne Zika virus has cast doubt on the accuracy of this categorisation.

1.2: The global impact of dengue virus

Dengue virus (DENV) is the causative agent for dengue disease; a febrile illness known for having both exanthematic and haemorrhagic manifestations that is now classified as either dengue without, or with, warning signs or severe dengue based upon the symptoms present and their severity (described further in 1.4.1). Before the clinical outcomes of dengue disease were reclassified in 2009, dengue disease severity was classified instead as either DF, dengue haemorrhagic fever (DHF), or dengue shock syndrome (DSS). Prior to 1939, DF was seen as a relatively mild disease; mortality was very low, DHF was very uncommon, and epidemics were rare due to the small size of densely populated urban areas in affected countries (5). During World War II, infected troops moving through South East Asia and the Pacific spread new serotypes of DENV leading to the first hyperendemic DENV-regions. Increased rural-to-urban migration within developing countries around the tropics post-World War II led to high population numbers and densities within urban environments which permitted greater levels of transmission and more regular epidemics. The increased use of air travel and the rapid globalisation in the coming

decades led to previously-unseen serotypes of DENV to be introduced into naïve populations which had a greater effect than could be predicted; while a single circulating serotype of DENV caused a low level of DHF and mortality, multiple strains circulating had a much higher level of DHF and subsequent mortality. This effect has been well documented in a number of case studies such as the 1981 Cuban DENV-2 epidemic which caused an unprecedented level of mortality in a population that had only previously been exposed to DENV-1 (6). The severity of the epidemic also put an incredible financial strain on the country, costing ~$103 million in healthcare, aid, mosquito control, and public outreach (7). Figure 1 shows the growth of dengue disease over time based solely on reported figures (8), however due to inadequate reporting and asymptomatic infections, these numbers are not believed to truly reflect the impact of dengue disease during this time.

Figure 1:

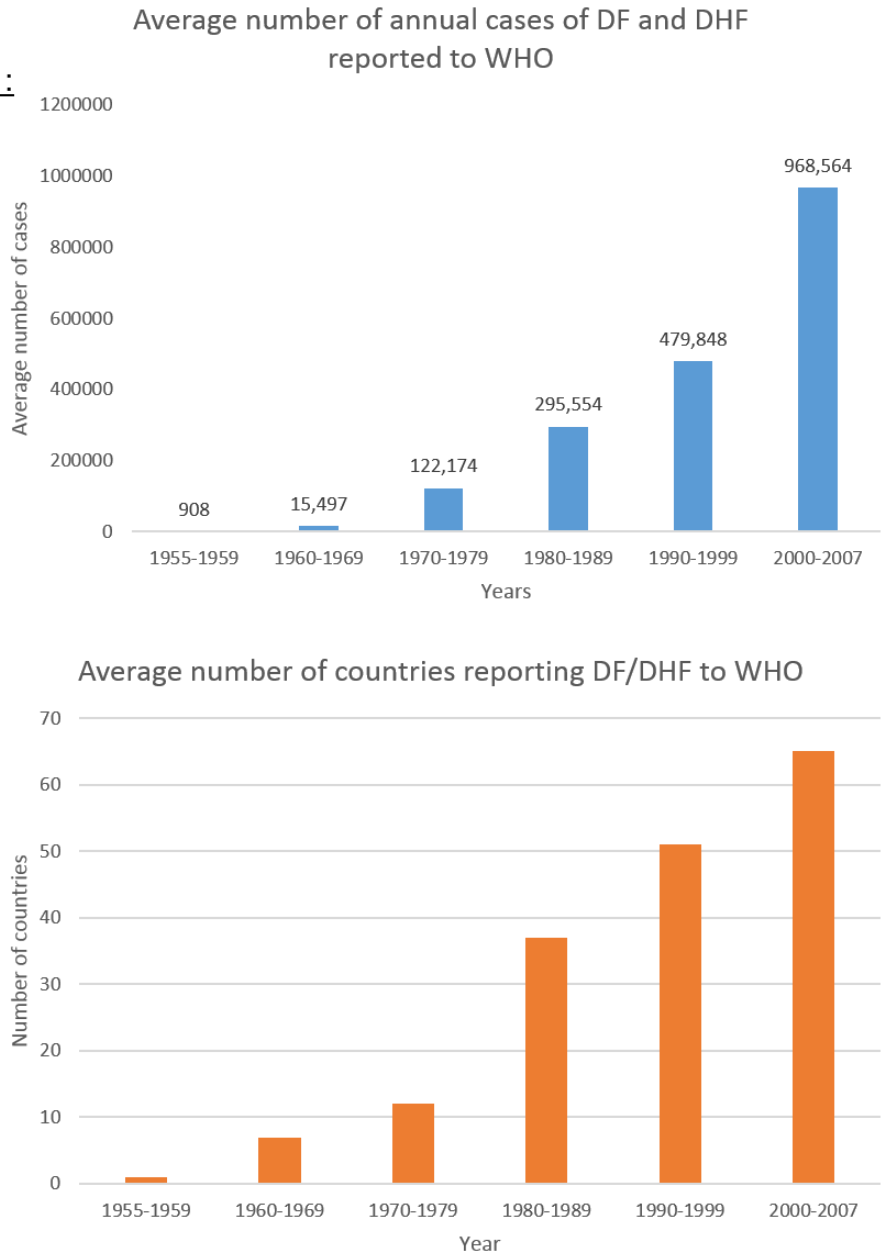

Figure 1: *The burden of dengue disease over time* - As reported to the WHO (8), the average number of annual cases of DF and DHF in each decade as well as the average number of countries reporting DF/DHF in those same decades

The current global infection rate is predicted to be ~390 million infections per year (95% - 294-528 million), 96 million of which are symptomatic (95% - 67-136 million) (9) and it is estimated that up to 3.97 billion people across 128 countries are at risk (10). Another study, which used a prediction of only 58.4 million symptomatic infections per year predicted a worldwide economic burden of $8.9 billion (95% - $3.7-$19.7 billion) (11) and, while recent advancements such as the licenced Dengvaxia vaccine suggest that those numbers could fall (12, 13), recent studies questioning the efficacy of the vaccine across young populations and in low-risk areas (14, 15) put into question whether those numbers will rise instead.

DF and DHF place a heavy strain on the healthcare systems of endemic countries with the World Health Organisation reporting ~500,000 hospitalisations and ~25,000 deaths per year due to the disease, most of these due to hypovolemic shock caused by high levels of haemorrhaging (DSS). Even for those who do not become hospitalised, the disease still has very important social and economic costs. The Disability Adjusted Life Scale (DALS) is a method for determining the impact of a disease not only on the emotional/physical level to the patient but also on a scale of loss of income and cost of medical care over the larger population. When analysed using this method, DF/DHF was ranked alongside serious, high-morbidity diseases such as tuberculosis and malaria (16).

DENV consists of 4 serotypes spread *via* the *Aedes aegypti* and *Aedes albopictus* mosquitoes which live predominantly within regions between the Tropic of Capricorn and the Tropic of Cancer. This means that autochthonous DF cases are prominent in equatorial countries which mostly consist of countries defined as ‘developing’. These countries lack the level of healthcare, infrastructure, and social/economic support for its population that is seen in developed countries meaning that a highly prevalent, highly debilitating disease can have a greater impact on the population of these countries than might be seen in more developed nations. Low cost, local methods for vector control have already been highlighted as a main strategy for countries in Latin America due, in part, to reasons like this (17).

1.3: The global impact of Zika virus

Zika virus (ZIKV) is the causative agent of ZF; a relatively mild febrile illness known mostly for the recent 2007-2017 Asian lineage pandemic which gained attention due to the appearance of neurological and developmental complications that were previously unseen in both the Asian and African lineages. First isolated in 1947 from a rhesus macaque from the Zika forest in Uganda (18), scientists quickly determined that ZIKV was also capable of human infection (19). However, it was established that ZF was a very mild disease; it had a roughly 80% asymptomatic rate, it had very mild symptoms even when symptomatic, and compared with other prevalent diseases in West Africa such as malaria, it appeared far less serious. For this reason, ZIKV research did not

advance very far in the next few decades however some miniscule surveillance did continue. In 1969 ZIKV was detected for the first time outside of Africa; in Malaysia (20), and then a few years later in 1977 it was detected in Indonesia (21). Although ZIKV exposure was found to be fairly common (around 40% of Nigerians tested in one 1970's study had signs of previous infection with ZIKV (22)) between ZIKV's isolation in 1947 and 2007 only 14 cases of natural ZIKV infection were ever confirmed (23).

Figure 2 shows the epidemiology of the outbreak that began in 2007 from its origin in Asia, across the Pacific, and finally into Latin America by 2015. ZIKV, like DENV, is transmitted *via* the *Aedes aegypti* and *Aedes albopictus* mosquitoes and while there were initial reports that *Culex spp* might be capable of transmission of the virus (24, 25), most studies so far have shown otherwise, at least in European/North American *Culex* species (26-28). However, unlike DENV, the virus is also capable of transmitting sexually (even after a vasectomy) (29), can form reservoirs in sexual organs (30), and has already shown it is a risk for contaminating the blood supply (31) and sperm banks (32) so the potential to spread to a wider geographic population than DENV is present.

Figure 2:

Figure 2: *The path that the recent ZIKV outbreak has taken along with the details of its evolution* -

1) 2007 - The Asian lineage of ZIKV broke out on Yap Island in Micronesia
2) 2012 – A descendent virus from the 2007 island broke out in the Philippines where the first recorded signs of neuropathogenicity were seen
3) 2013 – A related (likely descendent virus) then reached French Polynesia where the first cases of GBS and microcephaly were seen and high levels of asymptomatic infection led to an infiltration of the blood supplies
4) 2014 – The Cook Islands were infected next
5) 2014 – New Caledonia was also infected around the same time as the Cook Islands
6) 2014 – The virus reached Easter Island
7) 2015 – The neuropathogenic virus reached Latin America, appearing first in Natal, Brazil with the morbidities of GBS, ADEM, microcephaly, and other congenital neurological problems now associated with the virus outbreak (25)

Currently, it is very difficult to predict the global burden of ZIKV both economically and medically due to the high asymptomatic rate of ZIKV infection, a lack of global monitoring, and an

ongoing pandemic. However there are currently 59 countries reporting active transmission to the World Health Organisation and there have been predictions that ~2.17 billion people live in areas that potentially put them at risk of ZF (33). The U.S also made available $1.1 billion for ZIKV research/healthcare though a recent study has suggested that, depending on the successful transmission rates of ZIKV via mosquito bites, the cost could be far higher (34).

ZIKV has recently had a very large social impact across the globe with extensive media coverage (35) following the pandemic from as early as 2014-2015 when reports of microcephaly in congenital Zika syndrome (CZS) began to emerge (36). Since then, it has also been linked to Guillain-Barré syndrome (GBS) (37, 38) and acute disseminated encephalomyelitis (ADEM) (39) in adults as well as numerous further neurologic conditions in infants in cases of CZS (40). These have also been broadcast globally (41-43) causing travel fears (44) and loss of income *via* tourism (45) for affected developing countries.

1.4: Presentation of disease:

1.4.1: Dengue:

Dengue, as previously mentioned, is a febrile illness that has a wide spectrum of symptoms that can occur. These symptoms usually manifest between day 4 and day 9 post-infection but have been known to extend up until day 12. Rarely do they all manifest together but usually the common symptoms of fever and malaise are added to depending on severity of the infection with some symptoms appearing more commonly than others. Figure 3 shows the general spectrum of disease that is shown when infection with DENV occurs

Figure 3:

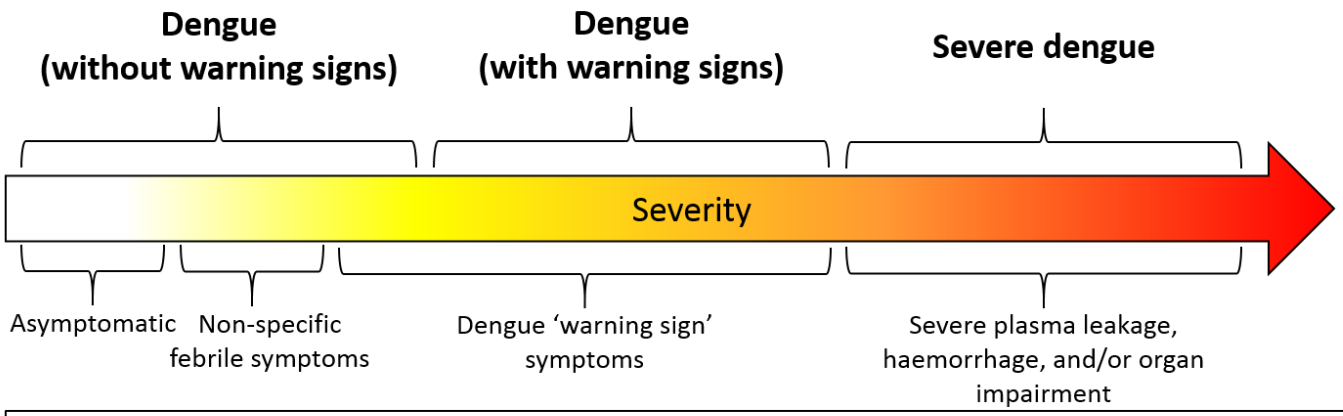

Figure 3: *The scale of DENV infection symptoms as ranked by severity of disease* – DENV infection causes a spectrum of disease and while most lie at the less severe end of symptoms, some do develop more life-threatening manifestations.

As shown in Figure 3, there are a number of general classifications for where certain symptoms would classify a patient on the scale of DENV disease. Further details are:

Asymptomatic – Asymptomatic infections are difficult to track and so the exact asymptomatic rate is difficult to estimate; current predictions though tend to range between 14-~50%. It used to be believed that the level of viremia in people with asymptomatic infections was not high enough to generate symptoms or spread the disease but recent research shows that asymptomatic carriers can pass the virus onto feeding mosquitoes (46).

Non-specific febrile symptoms – These symptoms tend to solely be malaise and fever making diagnosis of DENV infection extremely difficult as it is indistinguishable from many viral infections here. This level of disease is common in children who are experiencing their first DENV infection in an endemic country and due to the mildness of the fever, recovery often occurs naturally and without need for medical care within a week of the first symptoms appearing.

Dengue (without warning signs) – Dengue without warning signs consists of 'probable' cases of dengue which covers cases of fever in patients in dengue endemic areas that also have two other symptoms out of: nausea, macropapular rash, aching, leukopenia, or a positive tourniquet test. Although far less likely to progress into severe dengue, patients without warning signs can still develop severe dengue.

Dengue (with warning signs) – Dengue with warning signs is so named due to the increased severity of the symptoms outlining a great probability of progressing to severe dengue. Haemorrhagic manifestations of dengue are often small such as petechiae on the lower extremities, gingival bleeding, increased bruising, and even cases of thrombocytopenia but can extend to gastrointestinal haemorrhaging and systemic internal bleeding. The current list of 'warning sign' symptoms however is: mucosal bleeding, abdominal pain, persistent vomiting, fluid accumulation, lethargy, liver enlargement by >2cm, or a decrease in platelet count that causes an increase in hematocrit levels.

Severe dengue – Severe dengue patients develop more serious haemorrhagic symptoms along with severe vascular leakage into the pleural and abdominal cavities in what is known as 'the critical phase'. During this stage, the severe plasma leakage can lead to respiratory distress due to fluid accumulation around the lungs as well as other organ dysfunction like increased cardiac stress and liver damage. If the haemoconcentration is managed appropriately with IV fluid replacement then the chance of entering shock is greatly reduced. Shock usually occurs when a patient's fluid levels are not monitored or managed properly causing them to fall into hypovolemic/hypotensive shock. This causes further respiratory distress, cardiac arrhythmia, and organ failure. Roughly 25% of dengue-related shock patients die, rising to almost 100% if vascular collapse occurs. Once past the critical phase, patients then enter the 'convalescent phase' during recovery where they reabsorb the lost fluids from the abdominal and pleural cavity. Newer fluid

dosage guidelines have begun to reduce the problems of overcompensating for fluid loss, especially in children, which can cause hypervolemia, respiratory distress and death during the convalescent phase (47).

1.4.2: Zika

ZF has a much milder set of symptoms than dengue. Although the symptoms may be wide ranging, they rarely pose any risk to the patient. Approximately 80% of ZF is asymptomatic and although death can occur, it is incredibly rare and only usually occurs in those who are already severely immunocompromised. The common symptoms during symptomatic ZIKV infection consist of; fever, arthralgia, descending maculopapular rash, malaise, and occasionally abdominal pain and diarrhoea. The symptoms are very similar to other viral infections and other more common arboviruses in endemic areas such as chikungunya virus and DENV. Conjunctivitis tends to be more common with ZF than with dengue or chikungunya infection as well as limb oedema. Also hepatomegaly, leukopenia, and thrombocytopenia are far less common in ZF than in dengue.

ZF can have serious complications though with GBS being one of the most documented of ZIKV-caused maladies. GBS is an autoimmune condition targeting the myelin sheath of the peripheral nervous system causing the myelin to be degraded and removed from the axon of the nerve. Myelin, when produced and wrapped around the nerve axon by Schwann cells, allows for action potentials to pass more quickly along the nerve by increasing the maximum effective distance between each set of ion channels to efficiently polarise the axon (48). Exposed nerves lose efficacy and speed in transmitting action potentials leading to a tingling sensation in the limbs and muscle weakness and, if left untreated, nerve degeneration, atrophy, and therefore permanent paralysis (49). As this condition affects any peripheral nerves, breathing problems are relatively common and, depending on the speed of treatment, permanent nerve damage can occur. ZIKV is thought to cause GBS by a cell-mediated response to molecular mimicry (38) due to the speed of onset post-infection.

ADEM is also now associated with ZIKV infection and, unlike GBS, it is a demyelinating autoimmune condition that affects the central nervous system (CNS) rather than the peripheral nervous system (PNS). Symptoms have been described as similar to an acute attack of multiple sclerosis (seizures, headaches, ataxia, optic neuritis) however ADEM is a curable, single-attack disease driven by class switched IgG and triggered by trauma/infection whereas MS is an incurable condition driven predominantly by serum IgM but where individual attacks are stimulated through the same triggers. Like GBS, the demyelination can cause permanent damage if the autoimmunity is left untreated. ADEM and GBS have already been linked as a complication in numerous other viral infections (39, 50-53).

ZIKV presents its greatest and most specific risk when pregnant women become infected, especially if they are within the first trimester of pregnancy. This is because ZIKV is capable of infecting neurosphere cells and neural progenitors and then promoting apoptosis in them (54) and during the first trimester is when the brain is developing these key tissues. Retarded growth during this stage leads to a much smaller brain size which then presents as microcephaly when the infant is born. CZS at later stages can still affect other areas of the foetus' brain development due to the virus' ability to infect a wide variety of cell types, leading to other brain malformations such as calcifications, hypoplasia, ocular abnormalities, ventriculomegaly, enlarged extra-axial spaces, hypodensity of the white matter, and even reports of encephalomalacia in rare cases (40). It should also be noted that ZF has been linked recently to heart conditions like pre-eclampsia, palpitations, and arrhythmia though is not the only flavivirus to do so (55).

1.5: Flavivirus physiology, life cycle, and pathogenesis

1.5.1: Flavivirus genome organisation and physiology

The mature flavivirus virion consists of three structural proteins and a plasma membrane enveloping the genomic RNA. As shown in Figure 4, the RNA is encased within capsid (C) proteins which are in turn surrounded by a plasma membrane taken from the endoplasmic reticulum (ER) of infected cells. Embedded in the membrane is the membrane (M) protein - cleaved from the pre-membrane (prM) protein by a host furin or furin-like protease, and the envelope (E) protein which covers the M protein if it has been properly cleaved. The prM protein is not always cleaved during egress of the virus particle from infected cells, leading to totally-uncleaved and partially cleaved virion particles that have a 'rough' appearance to their surface as the prM proteins block the E proteins from forming their usual structure. These are known as immature/partially mature particles. There are minor differences in the macro structure of DENV and ZIKV as identified by cryo-EM (56, 57) but the principle of the protein organisation is the same.

Figure 4:

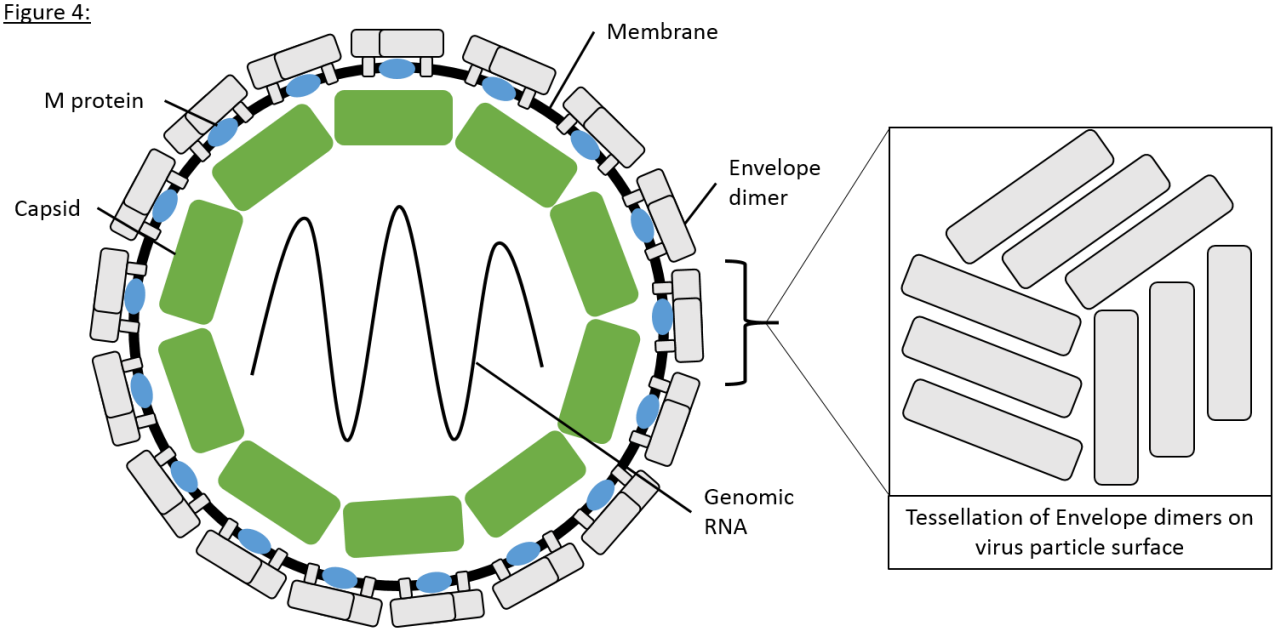

Figure 4: *The structure of a flavivirus particle* – The placement of the flaviviral structural proteins and genomic RNA in a fully mature virion particle is shown as well as the tessellation of the envelope protein dimer trimers that make up the outer surface of the virus

Along with structural similarities, flaviviruses also share the same general genome organisation and mechanism of polyprotein processing. The flavivirus genome is a non-segmented, positive-sense, single stranded RNA of ~11kb with a type 1 cap and a single open reading frame (ORF) which is translated into a polyprotein of ~3400 amino acids (aa). The polyprotein is both co- and post-translationally cleaved by both viral and host proteases into 10 proteins; C, prM, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 as shown in Figure 5. prM is then cleaved to M at the golgi body.

Figure 5:

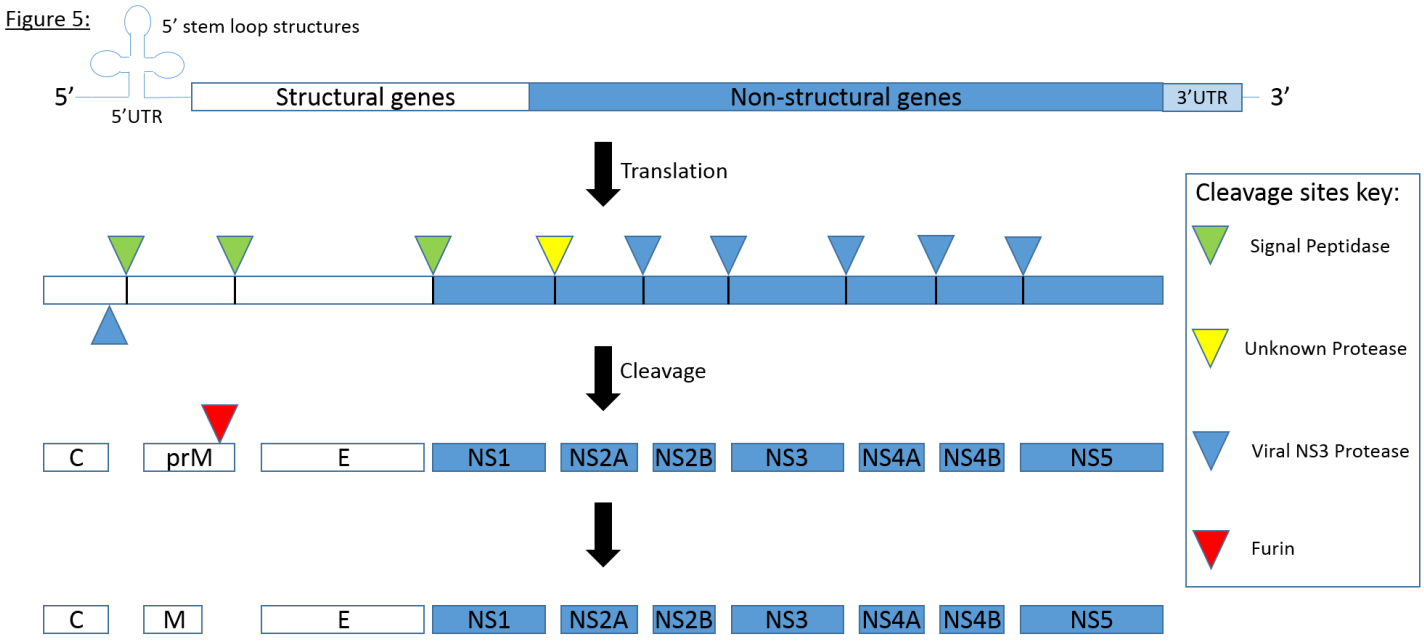

Figure 5: *Flavivirus genome organisation and polyprotein* processing - The translation of the flavivirus polyprotein and subsequent cleavage is shown including which proteases cleave which sites. Furin is a protein convertase located in the trans golgi and so prM is only cleaved to M just before the release of the mature virion to prevent premature membrane fusion. The other sites are cleaved during, and just after, translation.

The ORF is bordered by two untranslated regions (UTRs); the 5′ UTR and the 3′ UTR. These play important roles in the virus lifecycle with both the 3′ and 5′ UTR forming secondary RNA structures called stem loops that help in RNA-protein interactions with both viral proteins and host ribosomes and RNA-RNA interactions (58). The 5′ end of the genome is protected with a m^7GpppN-cap and, unlike some viruses, the 3′ UTR of flaviviruses is not poly-adenylated.

1.5.2: Transmission of dengue virus and Zika virus

Both DENV and ZIKV have a sylvatic and anthroponotic cycle as shown in Figures 6 and 7. It is likely that DENV originated from non-human primates as it is known to infect primates without the occurrence of DHF in either laboratory or natural infections (59). Humans would therefore be the 'spillover' hosts which would explain the severity of symptoms shown compared to non-human primates. ZIKV's species of origin is still currently unknown, however rhesus monkeys have been shown to be permissive for ZIKV replication and ZIKV was first isolated from a rhesus macaque. Upon ZIKV infection, rhesus monkeys do however show symptoms like human ZF and so are potentially another 'spillover' host. Antibodies against ZIKV infection have been found in a wide range of species as well such as: horses, cows, ducks, sheep, bats, and orangutans (60).

Both viruses cause viremia in a naïve host (as detected by PCR) for ~5-7 days during which they are infectious to feeding mosquitoes. Although little is known about the associations with ZIKV, faster passage from the midgut to the salivary glands of the mosquito has been linked with more severe dengue disease (61). For both viruses though, the average time it takes is ~12 days. Once in the salivary glands, the virus can be passed down the proboscis of the mosquito when it feeds and into the subcutaneous layers of the host skin.

ZIKV can also be transmitted sexually with detectable levels of ZIKV RNA being present in semen up to 125 days post-infection (62) and also in urine for ~40 days. ZIKV has been noted to pass from male to female (63) as well as through men who have sex with men (MSM) (64) but due to the virus' presence in cervical/vaginal fluid it has also been discovered to be capable of female to male transmission (65). The R0 of sexual transmission in ZIKV infection is lower than 1 (66) and so cannot sustain an epidemic.

Figure 6:

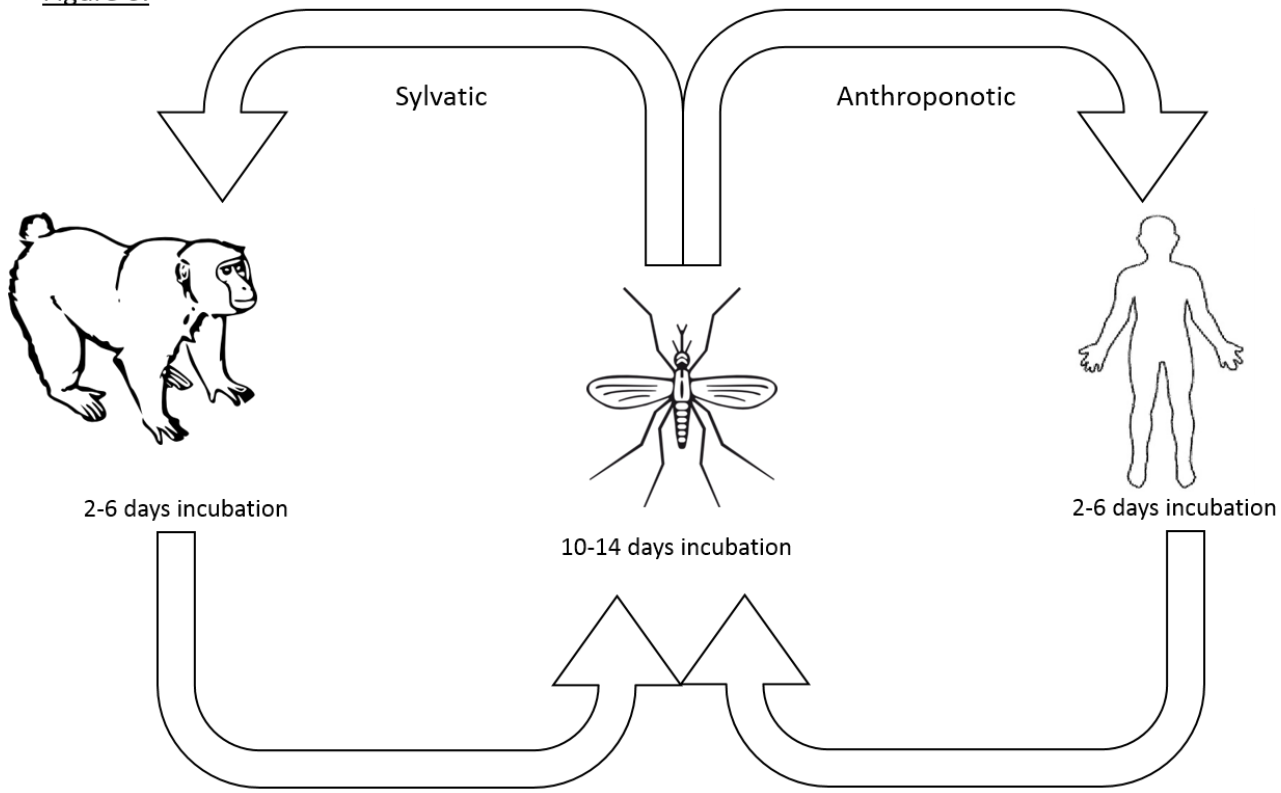

Figure 6: *The transmission cycle of DENV* - Other flaviviruses utilise other vectors such as ticks and other mosquitoes but DENV utilises *Aedes aegypti* and *Aedes albopictus* as their vector species. Although passage from a sylvatic cycle into an anthroponotic cycle and vice versa is possible, it is known to be a rare event in DENV due to the genetic differences seen in sylvatic cycle strains and anthroponotic cycle strains. It takes ~12 days from the virus to reach the salivary glands of the mosquito from the midgut and infected hosts are infectious for ~4-6 days

Figure 7:

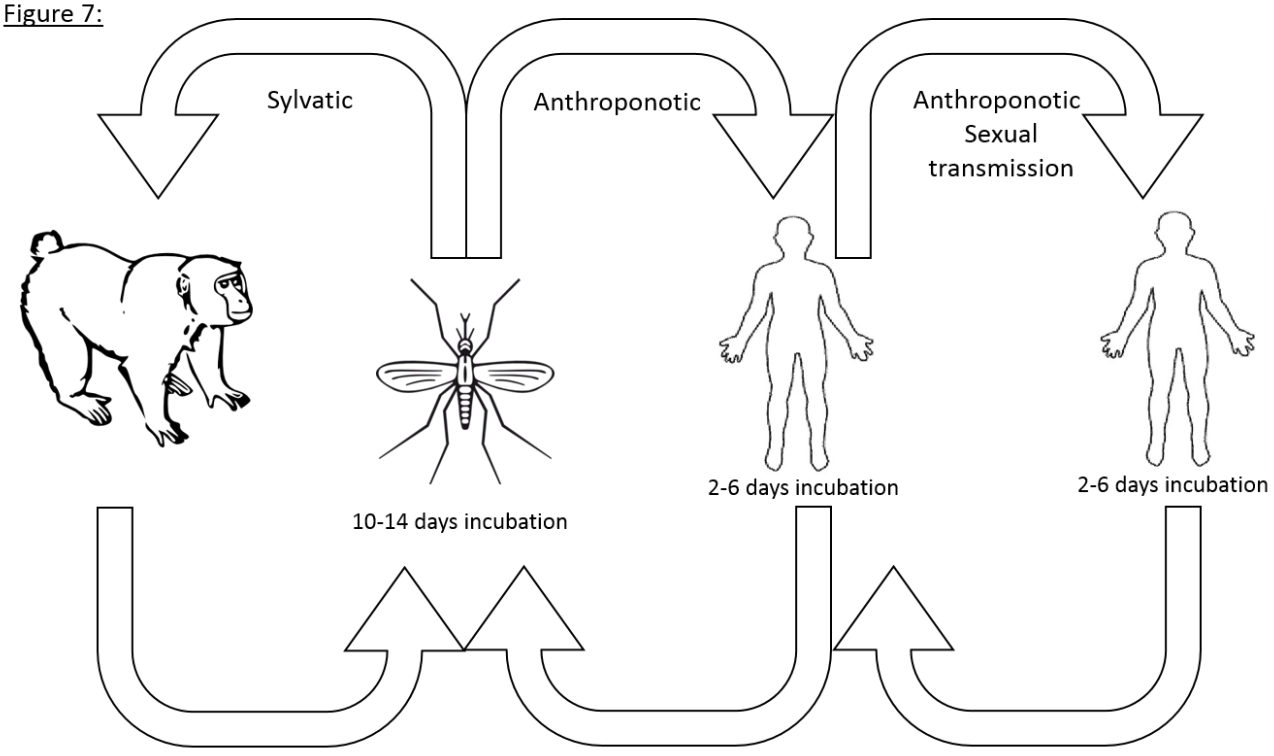

Figure 7: *The transmission cycle of ZIKV* – ZIKV also utilises *Aedes aegypti* and *Aedes albopictus* as their vector species. No research has so far shown how common the passage from sylvatic to anthroponotic cycles and vice versa is due to the lack of knowledge surrounding the virus within the animal reservoir. Unlike DENV, ZIKV can pass from human to human via sexual transmission also though not sustainably

1.5.3: Dengue lifecycle

Once DENV is introduced to the subcutaneous layer of the skin, its initial target is local immature Langerhans cells (67) but it can also infect macrophages and monocytes (68). These skin-specific dendritic cells are rich in receptors that are believed to be involved in DENV entry; dendritic cell-specific intracellular adhesion molecule-3-grabbing non-integrin (DC-SIGN), T-cell immunoglobulin mucin domain (TIM) receptors, and Tyro3, Axl, and Mer (TAM) receptors (69). As shown in Figure 8, DENV is taken up by clathrin-mediated endocytosis and proceeds through the normal endocytosis pathway.

Figure 8:

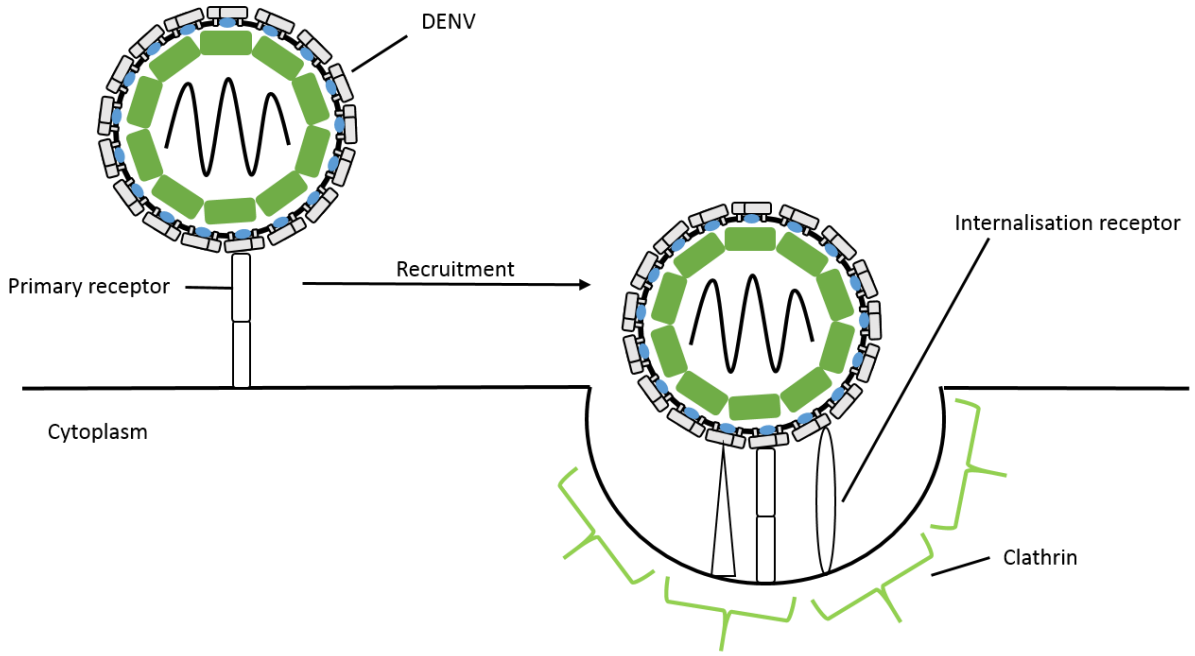

Figure 8: *DENV internalisation* - DENV particles bind to their primary/initial receptor (suspected to be DC-SIGN/TIM/TAM) before being recruited towards clathrin-endocytosis-rich sites. It is then bound by an internalisation/endocytosis receptor (the identity of which is currently unknown) which promotes its uptake by clathrin-mediated endocytosis

The endocytic particle then fuses with endosomal vesicles and progresses to an early and then late endosome where the pH drops to ~5.0-5.5. At this point, the DENV E protein undergoes a conformational change from its neutral pH conformation of a dimer to a low pH conformation of a trimer. In this conformation, the fusion loops are exposed and inserted into the endosome membrane to bring the membrane of the virus close enough to allow for membrane fusion (70). The capsid and the virus genome are then released into the cytoplasm. This process is shown in Figure 9.

Figure 9:

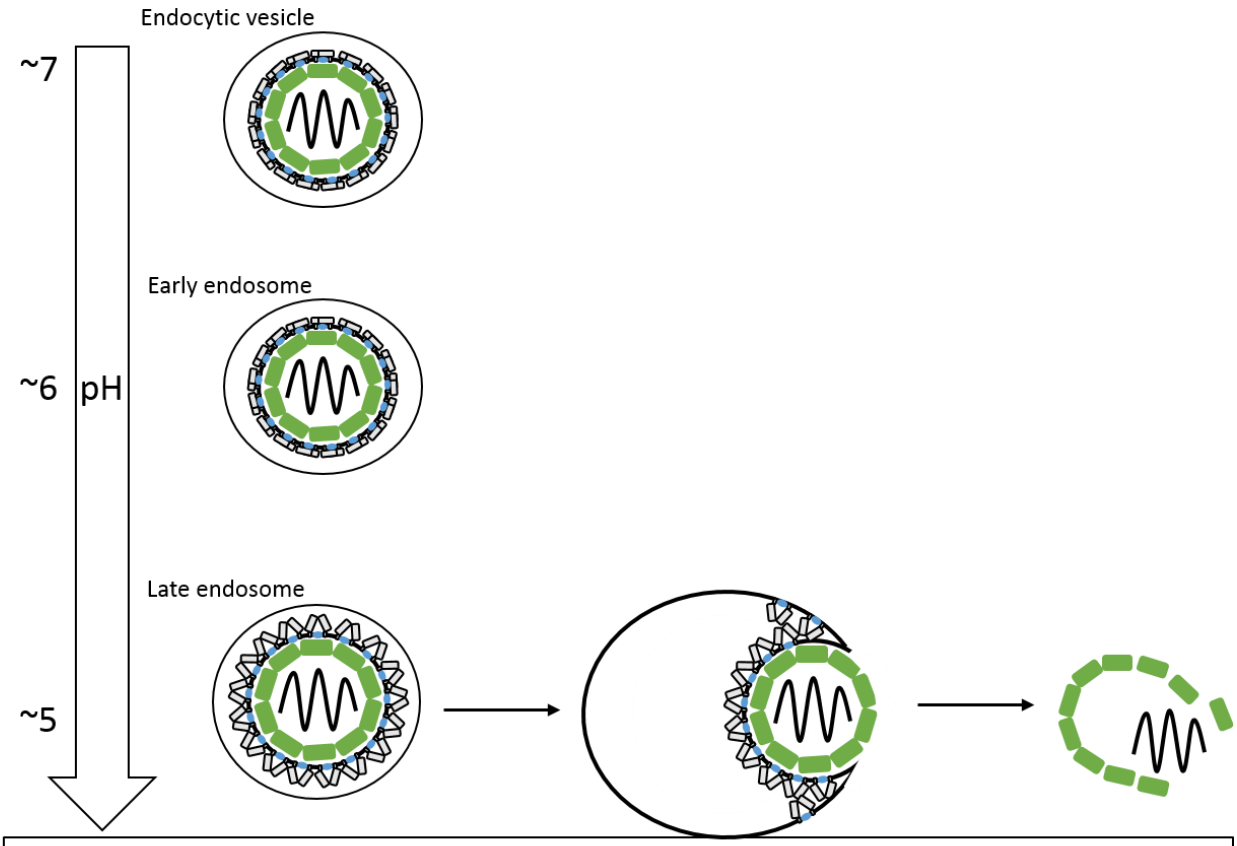

Figure 9: *DENV uncoating* - Clathrin quickly detaches from the endocytic vesicle which progresses through the early endosome and on to the late endosome where the low pH of 5.0-5.5 causes a conformational change in the E protein that pulls the viral membrane against the endosome membrane so that M protein can fuse them. Once this happens, the viral capsid is released into the cytoplasm where it too breaks apart and releases the RNA genome. The RNA then relocates to the ER.

As the genome is positive sense RNA, it can be translated by ribosomes without the need for the production of coding RNA. The cap structure helps load the RNA into the ribosome but the 3' UTR also plays a role in enhancing translation (71). The produced polyprotein is capable of autocleavage to release the non-structural proteins that then promote replication and modulate host responses. This processing is done within unstructured convoluted membranes induced by the virus.

Replication takes place within 'replication pockets' that are induced in the ER membrane (72) to produce a high concentration of viral RNA and viral replication proteins to increase the efficiency of replication and also to 'hide' triggers like dsRNA replication intermediates and uncapped ssRNA from the intracellular antiviral proteins such as retinoic acid-inducible gene I (RIG-I) (73). Inside the replication pockets, the genome must cyclize for the production of the negative sense antigenome to be created (74) which is necessary to produce positive sense genomes which can then be used to increase viral protein production and as genomes to be packaged into new virions.

Viral genomes then become packaged within invaginations in the ER membrane that are caused by the build-up of viral structural proteins on the internal side of the ER membrane. These bud off into the ER and then are transported in vesicles to the golgi to be matured by furin before being exocytosed. Figure 10 shows the steps of the virus lifecycle at the ER.

Figure 10:

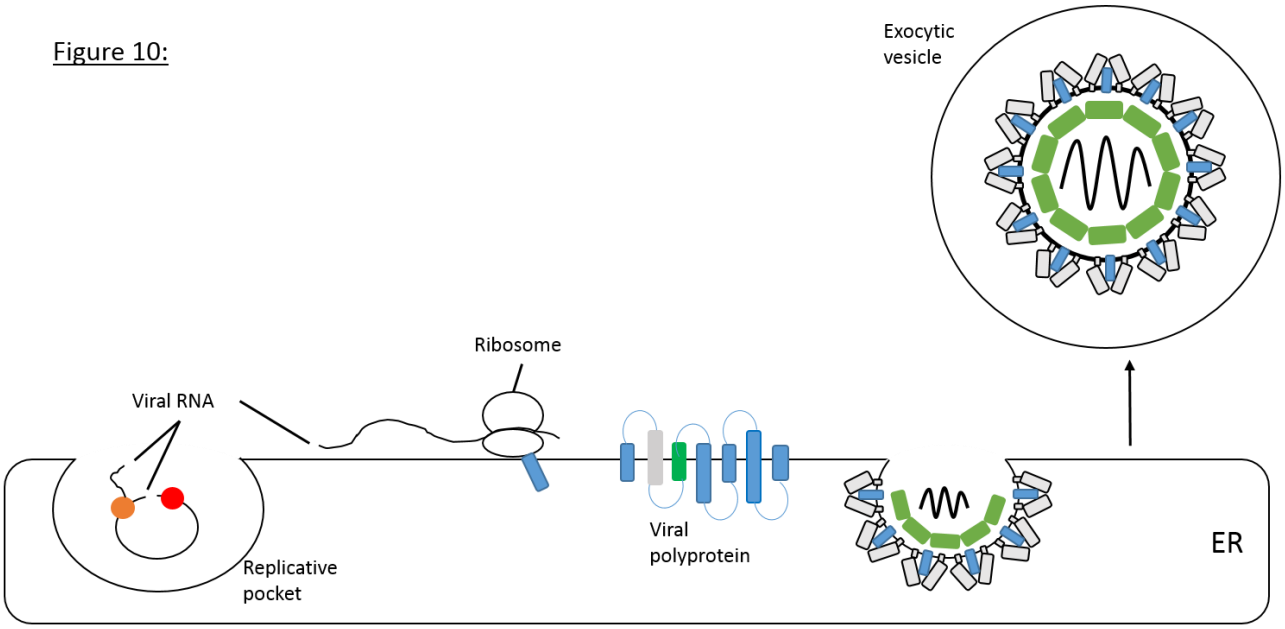

Figure 10: *DENV replication* - The RNA is processed through ribosomes using the stem loop structures in its 5'UTR to recruit it. Autocleavage of the polyprotein then generates the non-structural proteins needed to replicate the genome inside replicative pockets and manipulate the host cell's responses. After an eclipse phase of ~4 hours, virus particles are shed out from an infected cell with the virus budding into the ER to acquire a membrane coated in viral proteins before being exocytosed out of the cell via the golgi. At the golgi, furin will cleave prM into M protein to create a fully mature flavivirus. Without the maturation step, the virus cannot fuse membranes.

As the infection of DENV will have triggered the immature Langerhans cell to migrate to the local lymph node to mature, the virus will emerge from the eclipse phase and be capable of infecting surrounding dendritic cells, monocytes, and macrophages as well as escape into the blood from the lymphatic system and disseminate into other organs like the liver. There, DENV is capable of infecting not only hepatocytes but also Kupffer cells (75), using the abundant 37kDa/67kDa high affinity laminin receptor to enter hepatocytes (76). The immune response surrounding the liver is always fairly strong and damage to the liver during DENV infection is often associated with DENV-mediated apoptosis and immune response-mediated cell death. Cytotoxic CD8+ T-cells are implicated as playing a role in dengue-associate liver damage due to their prevalence in viral infection (77). Elevated aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels are seen in the majority of dengue with warning signs and severe dengue cases and while some patients show no liver symptoms, those that do can have symptoms that range all the way up to acute liver failure. However, myeloid cells are believed to be the primary target of DENV and upon infection they secrete high levels of cytokines and leukotrienes which have been shown to induce increased permeability in endothelial cell monolayers (78).

These are also believed to be the main source of vascular permeability as the appearance of DENV-specific CD8+ T-cells almost exclusively after haemoconcentration means CD8+ T-cell-mediated damage to the endothelium does not play an integral role in vascular leakage (79).

So far, no clear biochemical markers have been identified for distinguishing between dengue and severe dengue in the early stages of the disease and no markers have been found for determining whether particular genotypes of dengue virus cause a higher level of cytokine production. The cytokine storm generated predominantly by infected monocytes/macrophages is associated with the vascular leakage in dengue disease due to cytokines such as TNF-α and IL-1β increasing vascular permeability and errant platelet coagulation damaging the vasculature. However, NS1 protein binding and subsequent interference with the glycocalyx layer that provides barrier function in microvessels as well as the induction of complement by NS1 leading to the formation of terminal complement complex C5b-C9 have been shown to induce higher levels of vascular leakage. In gram negative bacteria, lipopolysaccharide (LPS) is well documented to bind toll-like receptor 4 (TLR4) to trigger a large cytokine response. DENV NS1 has recently been discovered to bind TLR4 and cause *in vitro* cytokine release and endothelial monolayer disruption (80) though its role *in vivo* is not yet known. It does however present a potential for further explanation as to the similarities between LPS-mediated endotoxic shock and severe dengue-associated shock. The similarities go further with greater numbers of circulating endothelial cells in a mouse model of dengue being shown to lead to higher levels of inducible nitric oxide synthase (iNOS) and peroxinitrate which caused greater apoptosis of cells and subsequent haemorrhage. In sepsis patients, higher numbers of circulating endothelial cells is linked with more severe septic manifestations and a worse prognosis as well (81).

1.5.4: Pathogenesis of Zika virus

The route that ZIKV takes through the human body is not well understood due to the only recent emphasis on ZIKV research. However, it is already known that ZIKV's main route of entry is the same as DENV; subcutaneous introduction *via Aedes aegypti* or *Aedes albopictus* mosquitoes, however ZIKV is also capable of sexual transmission. Currently, the primary target cell/organ of ZIKV is not known but ZIKV has been shown to be capable of infecting a large number of cell types (82, 83) including DCs (84) and macrophages (85) similar to DENV. Due to the high number of permissible cell lines, the cell entry receptor for ZIKV has been predicted to be either a very common cell surface receptor or, potentially, multiple receptors that can all be utilised as primary receptors by ZIKV. Tyrosine-protein kinase receptor UFO (AXL) was a prime suspect due to the previous evidence suggesting it was involved in the entry of a number of enveloped viruses and it's present on the surface of a wide range of cell types however knock-out (KO) studies in mice showed that removing AXL did not stop viral entry (86). AXL is however still thought to be heavily

involved alongside DC-SIGN, tyrosine-protein kinase receptor TYRO3 (Tyro3), and TIM-1 (87) much like in DENV infection.

Mouse models of ZIKV are not particularly reliable as, much the same as in DENV mouse models, the pathogenesis and disease progression is not comparable to human infection. ZIKV mouse models see no degradation of signal transducer and activator of transcription 2 (STAT2) which is a very well-known feature of human ZIKV infection (88, 89). As such, ZIKV mouse models must be IFN -/- (90, 91) with immunocompetent mice being highly resistant to infection. ZIKV mouse models also showed severe testicular atrophy to the point of infertility (92) which has so far not been reported within human infection.

Rhesus and cynomolgus macaques appear to be higher quality animal models for ZIKV infection however come with greater ethical restrictions and greater costs than mouse models. They do however provide more reliable descriptions of pathogenesis. A study performed by Osuna *et al* (30) showed the levels of viremia in 10 different macaques as well as the presence/absence of virus particles and RNA within saliva, blood, urine, semen, vaginal secretions, and cerebrospinal fluid (CSF) at specified timepoints over 21 days. The study used a virus similar to the Asian strain currently circulating in South America and as well as studying the presence/absence of the virus and its RNA, the study also kept track of the immune response generated by the infection as it progressed and during a re-infection 45 days after the initial infection. ZIKV was found in the urine of the macaques as early as day 2 post-infection and disappeared at roughly the same time as the viremia was resolved although a few sporadic reappearances suggest the possibility of reservoir sites within the urinary system. Even at 28 days, ~50% of the animals still had detectable ZIKV within the saliva and those that showed the presence of ZIKV within semen also showed signs that it had formed a reservoir there that was shedding detectable levels beyond the 28 day mark, which matches human reports of ZIKV being detected in semen months after infection (93). ZIKV was also detected within 50% of the macaques inside the CSF at various timepoints however it was only the macaques that had higher than mean viremia levels suggesting that neuroinvasion may simply be a ‘spillover’ effect of high viremia levels rather than a targeted route. This also seemed to be the case with all other bodily fluids except saliva. The three animals with the highest levels of viremia had detectable ZIKV in their CSF, urine, saliva, blood, and semen (all three were males) whereas the lowest peak viremia (female) only showed ZIKV RNA within her blood and saliva. The six animals in between showed a mix of results – all had ZIKV detectable in blood and saliva but its presence/absence in their urine, semen, cervicovaginal lavage, and CSF varied. The detection within the semen was also similar to what was seen in human males as the virus was detected in higher numbers in the prostate and seminal vesicles where it is predicted to reservoir in humans (29).

The spillover theory of ZIKV infection is a convincing one due to the large number of cell types currently found to be permissive to ZIKV infection; immune cells (84, 85), placental cells (82, 83),

neuroprogenitor cells (94), ocular nerve cells (95), fibroblasts and keratinocytes (87) are but a few. Likely ZIKV has a main target organ/cell type but its use of a widely expressed entry receptor or multiple receptors means that high levels of viremia expose greater numbers of permissive cells to the virus leading to invasion of other tissues. During infection of a pregnant mother though, the very high permissibility of placental cells and in particular placental macrophages and Hofbauer cells makes vertical transmission very likely. Hofbauer cells have been seen to migrate into the brain tissue of infected foetuses and so introduce the virus to permissive neural progenitor cells. Due to the apoptotic nature of ZIKV infection, the loss of neural progenitor cells leads to a diminished brain size as well as malformations and incomplete development (96, 97). Congenital infections are also linked with spinal cord deformities too suggesting that the infection of the undeveloped CNS extends into the spine or at least affects the development of the spinal nerves (98).

1.5.5: Antibody-dependent Enhancement

Antibody-dependent enhancement (ADE) is the term used to describe a well-documented side effect of DENV infection that raises the chance of severe dengue 15-80 times if the patient has been exposed to another serotype of DENV. As mentioned in 1.2, DENV infection was historically not considered a particularly dangerous infection but the globalisation that occurred brought new serotypes into populations that had developed immunity to a different serotype, leading to widespread cases of DHF/DSS like in Cuba in 1981 (6) so that now hyperendemic countries face a huge problem with multiple endemic serotypes circulating at once.

During primary infection, an effective antibody repertoire is developed that, if faced with the same serotype again, would provide a protective immune response that would neutralize the virus by stopping the conformational change of E protein so that even when the pH drops in the phagolysosome, the virus cannot escape into the cytosol of the phagocytic cell. Figure 11 shows a representation of a DENV particle in a phagolysosome.

Figure 11:

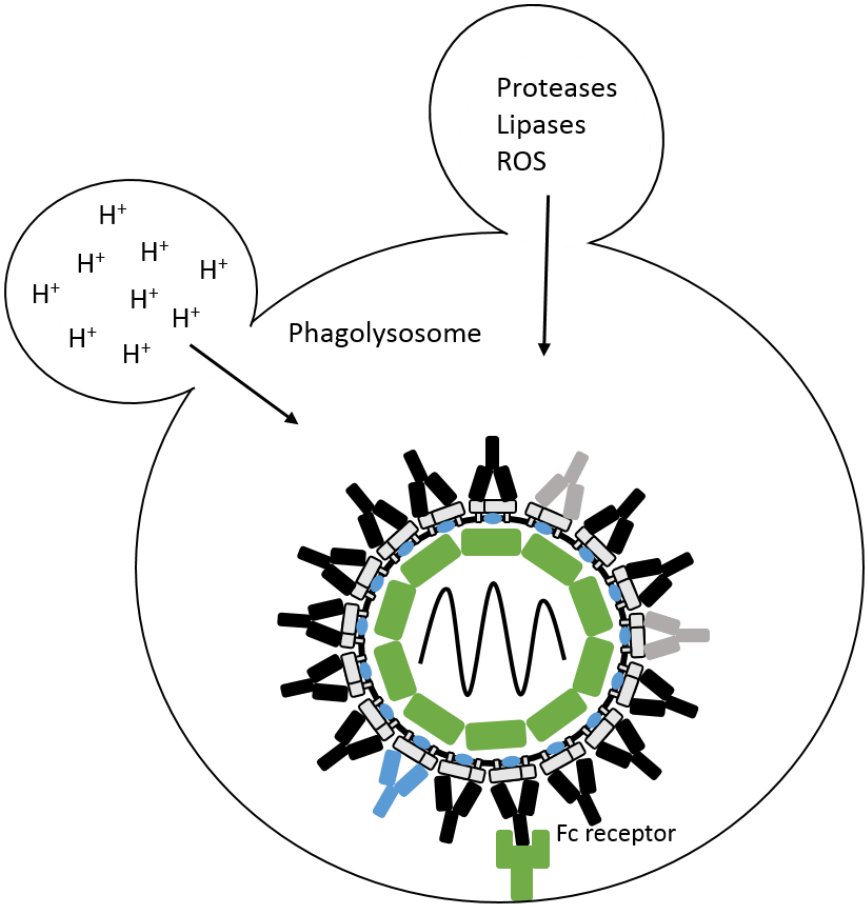

Figure 11: *Uptake of DENV with protective antibodies during primary infection* – A neutralised particle of DENV is unable to undergo the conformational change of E protein at low pH due to antibody binding. Taken into a phagolysosome via the Fc receptor, the virus cannot escape and is degraded by lysosomal enzymes

Occasionally during the initial phase of the secondary infection (with a different serotype), the antibody response to the primary infection is stimulated due to similarities in protein structure. DENV serotypes are ~70% similar to each other in amino acid sequence, giving a chance that an antibody epitope could partially recognise a resembling peptide on another serotype. However there are likely to be differences within that peptide as well, which would reduce the binding affinity of the antibody. The antibody response that was effective to the first serotype binds weakly to the second serotype and so fails to fully neutralize the virus. The virus is taken up by the Fc receptor as in a primary infection but unlike in a primary infection, when the phagolysosome pH drops, some of the E proteins can still undergo their trimerization and conformational change to induce membrane fusion and viral release (Figure 12). Even if there were antibody epitopes that are high-affinity binders to the second serotype present within the generated antibody repertoire, they would be unlikely to have been high-affinity binders of the initial serotype and so would not be produced at a high enough concentration to be considered neutralising (59). As the Fc receptor is faster at uptake than DENV's usual cell entry receptors, partially neutralised virus particles can infect at a faster rate than non-neutralised virus particles, meaning that an ineffective antibody response is more detrimental than no prior immunity at all as it leads to greater infection and higher viral load in early infection. As macrophages and monocytes are the primary target for DENV (as described in

1.5.3), this enhanced infection then leads to greater cytokine production, increased vascular permeability, and more severe symptoms.

Figure 12:

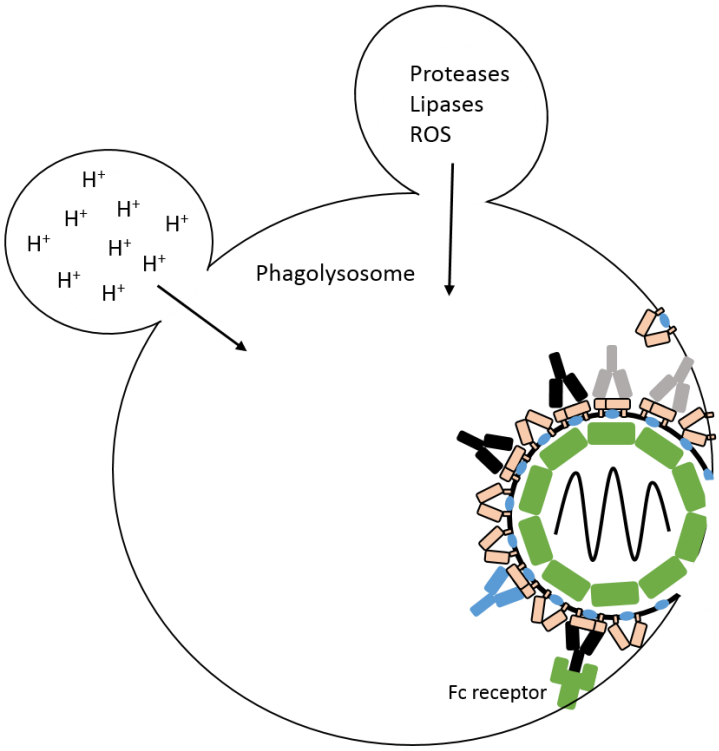

Figure 12: *Uptake of a second serotype of DENV with a non-protective, enhancing antibody response* – When a second serotype challenges the immune system, antibodies raised from the previous serotype's infection can be stimulated. Due to ineffective binding, the virus particle is not properly coated and can escape the phagolysosome once the pH drops, infecting the immune cell that has phagocytosed it

Cells in hypoxic conditions have been shown to require a higher titre of neutralising antibodies to neutralise viral particles as well as having a greater rate of uptake via the Fc receptor meaning that immune cells in hypoxic tissues have a far greater risk of ADE (99). Although this study was done *in vitro*, it reveals that the *in vivo* implications could be much more severe than previous *in vitro* ADE experiments have shown.

The two lineages of ZIKV have shown no signs of ADE between each other. In fact, antibodies raised against one serotype have been shown to be protective to the other serotype (100, 101). However, there have been reported cases of ADE between ZIKV and DENV (102). One study in particular presented evidence that all of its DENV-immune serum samples showed ADE for subsequent ZIKV infection that was not seen in DENV-naïve serum (103). Sharing the same vectors, the emerging ZIKV epidemic is passing through populations in DENV-endemic and –hyperendemic countries which is quite possible amplifying the ZIKV infection. There are currently no studies reporting whether ZIKV antibodies would increase the chances of ADE during DENV infection but there is potential for the current ZIKV epidemic to have another lasting effect by putting people at greater risk of ADE and subsequently more severe disease. Due to this risk, DENV ADE is now being taken into consideration in the development of ZIKV vaccines (104).

1.6: Flavivirus Proteins

1.6.1: Capsid

The flavivirus C protein is responsible for forming a protective capsule structure around the genomic RNA inside virus particles. Each immature monomer is ~114 aa in length and is cleaved by the NS2B-NS3 protease on the cytoplasmic side of the anchor peptide which creates the mature form of C protein (~100aa (~12kDa)) which then stabilizes in solution as a dimer. The host signal peptidase cleaves from within the ER lumen at the other end of the anchor peptide to release prM from the anchor peptide however this first requires the C protein-anchor cleavage by NS2B-NS3 protease so that C protein cleavage is a prerequisite for prM cleavage. The C protein dimers bind together in the presence of RNA to create an amorphic, physical protective barrier around the genome that is then enveloped in host membrane during viral particle assembly. Upon membrane fusion inside the lysosome of an infected cell, the capsid structure is released into the cytoplasm where the multimeric structure is believed to break apart to release the genome to be transported to the ER. Currently the process of encapsulation and capsid assembly, as well as capsid deconstruction upon virus entry are not well understood. However, what is known is that the C protein is not just a structural protein. A deletion study performed by Sansa *et al* (105) has shown that the DENV C protein can perform its structural role despite rather large deletions as long as two basic residue sequences at the N-terminus are intact but the infectivity of the deletion mutants which retained the ability to form a nucleocapsid was reduced, suggesting that there are other mechanisms performed by the C protein that affects the pathogenicity of the virus which was interrupted by the deletions (106, 107). The research into these other mechanisms is not yet extensive but there have been some breakthroughs. For DENV it was shown that the C protein interacts with lipid droplets early in infection which allowed for greater replication than if this action was interrupted with two L→S mutations within the α2 helices that disrupted the hydrophobic binding (108).

1.6.2: Pre-membrane/Membrane

Flavivirus prM protein is a ~166aa protein that is embedded in the virus lipid membrane by its C-terminal end, with the N-terminal end outside of the virus particle (109). prM is important for correct folding of the E protein during virus assembly, initially forming a heterodimer with the E protein. The N-terminal end of the protein prevents the E protein from structural rearrangements due to premature fusion during the process of virus assembly and secretion. Once the prM protein is cleaved by the host protease furin at the golgi body, the E protein rearranges on the virus surface to form a E protein homodimer, resulting in the mature virus particle. The prM protein is

considered matured and is then referred to as the membrane (M) protein. In the mature virus particle, the E protein dimers cover the M protein fully. Partially-mature and immature virus particles are produced when the prM proteins that coat the virus are either or only partially cleaved or not at all. Immature virus particles are not infectious but may be taken into cells by the route of ADE by antibodies recognising the prM protein (110). There has also been evidence to suggest that immature particles can be infectious in immature DCs via interaction with DC-SIGN (111).

1.6.3: Envelope

E protein is a ~495 aa surface glycoprotein anchored to the viral envelope membrane by two transmembrane helices. It is fairly well conserved across flaviviruses with a ~40% similarity at the aa level (112). There are 180 E proteins per virion consisting of 90 pairs of E protein homodimers that interact with 2 M proteins each and are arranged on the virus surface in a herringbone pattern (113). The E/prM interaction occurs within minutes of translation and processing of the polyprotein and involves two conserved helices located towards the C-terminal end of the protein (114); in DENV-2 these equate to aa 398-420 and aa 426-448. Each E protein is made up of three domains; domain I which is the central structural domain, domain II which is the dimerisation domain, and domain III which is the Ig-like receptor binding domain.

The E protein's role is to act as a cell surface binding and fusion protein. It interacts with cell surface receptors using the glycosylated domain III. The glycosylation patterns vary even from strain to strain which can affect the affinity of different receptors to be utilised for cell entry and this can play a role in the virulence of the virus. For example, the YFV-17D vaccine strain is largely attenuated by mutations within E protein domain III so that it has a very high affinity for glycosaminoglycans (GAGs) which are expressed on the surface of many cells as well as being very prevalent within the extracellular matrix (ECM). This means that YFV-17D does not remain in the blood for long once it has been introduced and often becomes harmlessly bound to the ECM. Even cells that are infected are believed to be re-infected by progeny virus that attach immediately after being secreted from the cell, leading to an ineffective infection that is easily overcome by the immune system (115). After the uptake of the virus, the phagolysosome acidifies and this acidification causes the E protein dimers to dissociate and then reform irreversibly as a trimer. The trimer inserts itself into the membrane of the phagolysosome and the E protein bend at the domain I/domain II hinge, fusing the envelope membrane with the phagolysosome membrane to release the capsid and its contents into the cytoplasm (112).

The E protein is also responsible, alongside prM protein, for the budding of the membrane that occurs in the creation of viral particles. The expression of prM and E alone causes the release of virus-like particles that are structurally and functionally like the flavivirus membrane (116). The E protein is the primary target of the host antibody response. Antibodies raised within mice to domain

III of E protein are often neutralising as they can obstruct the receptor binding region. As previously mentioned, glycosylation on the E protein varies between flaviviruses and even strains within those species, which means that those antibody responses towards E protein domain III are fairly specific. This fact has garnered some therapeutic interest because most flavivirus ELISA tests run with high levels of cross-reaction (117) that means clear identification is not possible without RT-PCR tests or viral plaque assays which are more costly and take more time. ELISAs based on domain III of E protein have been shown to identify the correct infection using a patient's serum with higher sensitivity than with larger viral antigens (118). Vaccination with WNV E domain III generated an antibody response that was protective *in vitro* and *in vivo* to WNV challenge but also partially protective to DENV (119). When JEV was incubated with the same serum and then introduced into mice, the development of brain injuries was prevented compared to untreated virus (120). This gives E protein domain III an enticing prospect of being used in vaccines. Other therapeutic uses for E protein domain III have included the generation of mAbs that could be infused as a treatment which was shown, as well as the use of domain III as a competitive binder of cell entry receptors, to reduce virus infection of cells (118).

1.6.4: NS1

NS1 is a highly conserved protein across flaviviruses. It is 352 aa long and has a molecular weight of 46-55 kDa depending on its glycosylation status. It is found commonly in one of three native states: a secreted hexameric lipoparticle form (121), a cell surface-bound dimeric form, or a virus-induced intracellular vesicle-bound form (122). NS1 forms a hydrophobic dimer without the assistance of other proteins (123) and associates with lipids *via* a glycosyl-phosphatidylinositol (GPI) linkage (124).

The secreted form is an early marker of disease, especially in primary infection, and it is also known for being highly immunogenic and potentially plays a role in flavivirus virulence. The hexamer is held together by weak hydrophobic bonds that are easily disrupted although this has no effect on secretion levels of NS1. When secreted, NS1 has a number of targets such as; the complement system (125, 126) and coagulative regulators (127), and has also been reported to accumulate within the late endosomes of hepatocytes *in vitro* (128) resulting in increased DENV virulence upon infection. The antibody response to NS1 also plays an important role in secondary infection though it is not yet clear what defines a protective response and what defines a detrimental one. Although the antibody responses are not neutralising, antibodies capable of fixing complement allow cell surface-bound NS1 to be targeted and infected cells to be destroyed *via* the membrane attack-complex (MAC). There has however been recent evidence to show that non-complement-binding antibodies can also provide protection by binding cell-surface NS1 and directing the immune response towards the infected cells *via* the Fc receptor (129). The antibody

response has been known to cause problems too. NS1/antibody immune complexes have been shown to cause vascular damage and antibodies raised against NS1 can be cross-reactive against the ECM and epithelial cells resulting in apoptosis of uninfected cells (130).

Non-secreted NS1 is required for flaviviral replication though its role is still speculative. It has been shown to co-localise with dsRNA and other flaviviral replication proteins but given its placement on the internal lumen of the ER (while the replication complex is external), it likely performs a scaffolding/recruitment function for the replication complex (131). WNV NS1 has also been shown to inhibit the TLR3 response in WNV replicon-expressing cells (132).

1.6.5: NS2A

NS2A is a hydrophobic protein of ~231 aa in length that is membrane-associated and, like many flaviviral proteins, has multiple roles *in vivo*. It is not highly conserved between different flaviviruses with the NS2A proteins of DENV and ZIKV only having a ~24.8% similarity at the aa level. This explains the differences in some of NS2A's function in different flavivirus. For example, the presence of ZIKV NS2A in radial glial cells was shown to disrupt adherens junctions and lead to premature differentiation into neurons both *in vitro* and *in vivo* as show by *in utero* electroporation transfecting the cortex of embryonic Crl:CD1 (ICR) mice with the ZIKV NS2A gene (133). This phenotype is seen in native infection of foetuses and so NS2A is believed to the main contributing factor behind this aspect of the neuropathogenesis seen in CZS. The same study tested DENV NS2A in the same manner and did not see the junction disruption or the premature differentiation. Many, if not all, flavivirus NS2A proteins play a role in modulating the IFN pathway, however they do so in different ways. NS2A from DENV-1, -2, and -4 inhibit the IFN-β induction pathway by interfering with tank binding kinase 1 (TBK1) phosphorylation of interferon regulatory transcription factor 3 (IRF3) in conjunction with NS4A (134). For WNV, NS2A is also essential for the inhibition of IFN-α/β induction (135, 136) providing evidence that NS2A interference with the IFN-signalling pathway could be a common interaction within flaviviruses.

NS2A does however retain a set of functions common to all flaviviruses. Its localisation within the replication complex is well-documented with its role predicted to be mostly structural in keeping the replication complex close to the membrane (137) likely *via* its strong interactions with RNA. It also plays a role in virus particle assembly. Mutagenesis of the YFV and WNV NS2A proteins resulted in the production of non-infectious prM/E particles but no C protein or RNA was loaded into them, leading to the prediction that NS2A may be involved in transporting and/or packaging viral RNA (138).

1.6.6: NS2B

The NS2B protein is a ~96 aa (~14 kDa) transmembrane protein consisting of two transmembrane domains (TMD) at the N- and C-termini of the protein with one central domain. The NS2B protein functions as an essential cofactor for the NS3 protease. The central region of NS2B expressed without the two TMDs has the capability to solubilize (though not activate) the NS3 protein (139). When interacting with the NS3 protease, the C-terminal end of the central region (aa 68-96) inserts itself within the substrate binding pocket using a hydrophobic β-hairpin structure and plays a direct role in protein-protein interactions with the substrate. Therefore, the NS3 protease cannot function without the NS2B cofactor (140). The TMDs of NS2B also play an important role by localising NS3 to the membrane *via* the NS2B/NS3 protease binding – a localisation that is important in efficient viral replication.

1.6.7: NS3

The NS3 protein is a ~619 aa (~69 kDa) multifunctional protein with the N-terminal domain containing a chymotrypsin-like serine protease, modulated by NS2B, which cleaves both *in trans* and *in cis* and the C-terminal domain containing a helicase domain which bears some structural similarity to the conserved RecA fold responsible for nucleotide binding and ATP hydrolysis. The protease domain (~167 aa) and its structural relationship with NS2B was discussed in 1.6.6. Its individual structure was described by Luo *et al* (139). The helicase domain (~446 aa) consists of three subdomains; subdomain 1 and 2 form a RecA-like structure that contains 8 highly conserved motifs that control RNA binding, ATP hydrolysis, and the mediation of the binding and activation of both sites, while subdomain 3 forms the RNA binding tunnel that feeds RNA into the RNA binding site (140). The ATPase site within NS3 helicase subdomains 1 and 2 is also responsible for RTPase activity. NS3 helicase subdomain 3 is also responsible for the interaction with NS5.

NS3 plays numerous highly important roles within the flavivirus lifecycle. The NS3 protease cleaves *in cis* and *in trans* to proteolytically process the flavivirus polyprotein; cleaving the C protein (with host signalase) from prM protein and also cleaving between NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5 (141). NS3 has also been shown to undergo autoproteolysis within mammalian cells though the biological relevance of this action is not yet understood (142). The NS3/NS5 heterodimer is a key flaviviral feature, acting as the core component of the replication complex. The helicase action of NS3 releases ssRNA from dsRNA to be used as either a template for replication in the case of antigenomes, or to be processed and packaged into viral particles in the case of newly formed genomes. The NS3 protein also processes the newly formed genomes using the RTPase activity of the helicase domain to trigger the first step in viral capping that is then completed by the NS5 protein.

NS3 also plays a role in the modulation of the host response to infection. A study analysing the human cellular interactome of the NS3 protein from multiple flaviviruses including; DENV, the WNV strain Kunjin (WNV_{Kun}), Japanese encephalitis virus (JEV), tick-borne encephalitis virus (TBEV), and Alkhurma virus (ALKV) revealed a tendency for NS3 to interact with 'hub' proteins allowing it to have a much more prominent effect on the regulation of host cell pathways (143). In human cells, though not in mouse cells, the DENV NS3 protease was observed to cleave stimulator of interferon genes (STING) (144), a regulator of both IFNα/IFNβ induction as well as cleaving protein arginine methyltransferase 5 (PRMT5) (143), a member of the mitogen-activated protein kinase (MAPK) signalling cascade, whilst the JEV NS3 protein cleaved AP-1 (145), a transcription factor important in the transcription of genes relating to inflammation and apoptosis. The DENV NS3 protein has been reported to co-localise and bind to fatty acid synthase (FASN) during infection (146). FASN is an important modulator of the lipid profile of DENV replication centres facilitating efficient replication.

1.6.8: NS4A

NS4A is a membrane-bound ~127 aa (~16 kDa) protein containing TMDs that induces curvature of the membrane that it is situated in. Within the viral polyprotein, it is located alongside a conserved 23 aa linker protein known as the 2K fragment. When the polyprotein is cleaved, the NS4A/2K boundary is cleaved by NS2B/NS3 protease which then triggers the cleavage of 2K/NS4B by a host signalase. Despite being cleaved off, the 2K fragment plays an important role in NS4A's function. In DENV infection, the 2K fragment prevents NS4A activity and must be cleaved free before it can function and, as shown by studies on WNV, the 2K fragment must be present and attached to NS4A before proteolytic processing or NS4A accumulates in the golgi and cannot have its native effect (147). When expressed and cleaved, mature NS4A can naturally form a heterodimer with NS4B as well as NS4A oligomers *via* TMD1 as well, though the role of NS4A oligomers is not yet known (148).

NS4A has been observed to co-localise with dsRNA and within flavivirus infection, it is known to form a key part of the replication complex. Its interaction with host vimentin and viral NS1/NS4B allows it to regulate and initiate the construction of the replication complex. The interaction with NS4B can only be performed by monomers of NS4A. The WNV NS4A protein has also been shown to regulate the ATPase activity of the NS3 helicase (147). Its capability to induce curvature within the membrane also contribute to the induction of the replication pockets that are seen within the ER and virally-induced membrane structures in flaviviral infection. NS4A is also responsible for triggering the formation of these virally-induced membrane structures (149, 150).

Outside of the replication of the virus and interaction with viral proteins, NS4A also interacts with host factors to assist in flaviviral infection. Autophagy is well documented to be used by

numerous flaviviruses either to protect from apoptosis or as a site of replication in the case of DENV-2. NS4A from DENV-2 and Modoc virus (MODV) induced an increase in cellular autophagy *via* the PI3K pathway (151). This response is thought to be important during infections such as DENV when post-acute-phase infection of hepatocytes and epithelial cells occurs, however the role of PI3K in preventing apoptosis has only been shown *in vitro* so far (152). Its presence can trigger the unfolded protein response (UPR) within the ER and can play a role in IFN suppression though that remains species and serotype specific. During DENV-1 infection, for example, NS4A strongly inhibits RIG-I/TBK –mediated IFNβ production, but that inhibition has not seen observed in studies using DENV-2 or DENV4 infected cells (153).

1.6.9: NS4B

NS4B is a highly hydrophobic protein of ~249 aa (~28 kDa) in size that is fairly well conserved at the aa level - ~35% across mosquito-borne flaviviruses and ~78% between all DENV serotypes. The high hydrophobicity of the protein has prevented structural characterisation of the protein at atomic level resolution so far, however there is a generally accepted topology model from biochemical studies of 3 TMDs and 2 membrane-associated domains (MADs) (154). The hydrophobicity is also one of the reasons the protein readily forms homodimers and heterodimers with NS4A without needing any additional proteins present (147, 155).

The formation of NS4A/NS4B heterodimers is clearly important for flavivirus replication as prevention of this interaction or a reduction in the affinity of this binding by mutation of the binding site leads to a greatly diminished capability to replicate. The interaction is also important for improving NS4A's ability to rearrange the ER. The dimerisation with NS4A isn't the only role NS4B plays within the replication complex with the 30 aa cytoplasmic loop of NS4B known to bind subdomains 2+3 of the NS3 helicase to increase the dissociation of the helicase with ssRNA and so improve replication efficiency (156). A single point mutation at aa 52 within a non-lethal strain of DENV increased lethality to 80% within mice (157) through changes in replication efficiency only. The reverse was also shown with a naturally lethal strain within mice showing how integral NS4B is to DENV replication.

When expressed alone, NS4B inhibits protein synthesis, reduces ER vesicle production, and – in DENV – stimulates the UPR. This reduces the production and release of IFN upon infection as well as providing more ER membrane for use in replication. NS4B's role in impeding the IFN pathways extends further though. For example, DENV NS4B inhibits the phosphorylation of STAT1 and while it hasn't been shown for other flaviviruses, both YFV and WNV NS4B proteins co-localise with STAT1 indicating they may also have a role in STAT1 signal transduction. Both DENV and WNV NS4B also inhibit RIG-I/mitochondrial antiviral-signalling protein (MAVS) signalling by

blocking TBK1/IRF3 phosphorylation and so prevent the induction of IFN. YFV NS4B also blocks RIG-I/IFN signalling but instead binds and sequesters STING (158).

1.6.10: NS5

NS5 is the largest flaviviral protein at ~899 aa (~103 kDa) and has been well conserved across *Flavivirus* species. It consists of two main enzymatic domains; an N-terminal methytransferase (MTase) domain and a C-terminal RNA-dependent RNA-polymerase (RdRp) domain. The two domains are critical for capping and replication of the flaviviral genome respectively. NS5 localises differently for different flaviviruses and occasionally for different serotypes within those species. DENV NS5 is well characterized to localise both in the nucleus and the cytoplasm but DENV-1 and DENV-4 have a predominantly cytoplasmic localisation whereas DENV-2 and DENV-3 have a much more prominent nuclear localisation (159). There are multiple sites found to be responsible for this variation in localisation between the serotypes; the βNLS, αβNLS, and a C-terminal 18aa region (160, 161). YFV , WNV, and JEV all have NS5 proteins that predominantly localise to the cytoplasm but a significant minority is also found within the nucleus (162). The localisation of NS5 within DENV serotypes shows that replication and IFN inhibition is not significantly altered when the localisation of the NS5 protein is changed (160). Therefore the role of NS5 within the nucleus must pertain to another function.

As the viral RdRp, NS5 is responsible for the production of genomes and anti-genomes from viral RNA. Recruitment of the NS5 protein to the 5′ end of the viral RNA is dictated by structural elements within the RNA that also control the cyclization of the genome for antigenome production (74). NS3 and NS5 make up the core of the replication complex with NS5 shown to be essential for NS3 NTPase activity (163). Although the polymerase has retained a high level of conservation across the flaviviruses there are differences between them. ZIKV heavily favours Mg^{2+} as a cofactor, requiring 5 times as much Mn^{2+} as DENV to function comparably (164).

N7-methylation is key in the processing and protection of the viral genome as one of the final steps in RNA capping. Methylating the guanine cap at the N7 position, and then subsequent 2'-O-methylation of the base attached to the m^7Gppp cap, gives rise to the classical cap-1 structure which is essential for the translation of the viral genome (165). The MTase activity of NS5 covers both N7-methylation and 2'-O-methylation with both using S-adenyl-L-methionine (SAM) as their methyl donor. 2'-O-methylation protects the virus from detection by the immune system by mimicking cellular mRNA and loss of this kind of protection, although not lethal, is highly attenuating to the virus (166). m^6A-methylation is also important in the lifecycle of some flaviviruses but this is only performed by host MTases (167).

Like the other non-structural proteins of flaviviruses, NS5 interacts with host proteins as well as with viral proteins/RNA. All NS5 proteins intervene in the IFN pathway by preventing transduction of the IFN-receptor signalling at various stages. DENV is well documented to degrade human STAT2 using the E3 ubiquitin-ligase UBR4 to target STAT2 to the proteasome (168). The inhibition of type-1 and type-3 IFN responses seen in ZIKV infection is indicative of a loss in STAT2 function (89). ZIKV binds STAT2 before phosphorylation and, like DENV, leads to proteasomal degradation but in a UBR4-independent manner (88). The expression of ZIKV NS5 (ZNS5) also leads to a ~70% reduction in interferon induced protein with tetratricopeptide repeats 1 (IFIT1) expression. These responses are specific to human hosts as wild-type (WT) mice are resistant to ZIKV infection and have to be IFN-deficient for mouse studies to take place. WNV NS5 inhibits the surface expression of IFNAR1, YFV NS5 binds phosphorylated STAT2 to stop it binding to interferon-stimulated response element (ISRE) promoters, LGTV NS5 interrupts interferon alpha and beta receptor subunit 1/2 (IFNAR1/2) signal transduction, and JEV NS5 blocks STAT1/Tyk2 phosphorylation (169).

Outside of IFN inhibition, NS5 still has important host interactions. In a Y2H study involving flavivirus NS5, common gene ontology (GO) groups that were enriched included RNA binding, cytoskeleton proteins, and intracellular transport with numerous examples published in the same study (143). A DENV-specific interaction study saw an enrichment of UPR proteins interacting with NS5 (170) as well as another study showing that DENV NS5 interacts with U5 snRNP spliceosome proteins to alter splicing patterns (171). It even utilises host factors to assist in replication with host cyclophilin A (CyPA) being shown to localise with NS5 and the replication complex and its enzymatic activity shown to be important in the replication of WNV, YFV, and DENV (172).

1.6.11: Similarities between ZIKV and DENV NS5

Outside of the basic roles of NS5 as an RdRp and MTase and despite some minor differences as discussed in 1.6.10, the ZIKV and DENV NS5 proteins have been shown to possess decent functional and structural similarities in a few studies which directly compared them. The structural similarities of the two NS5 proteins was compared by Baez *et al* (173) to examine whether, from a structural point of view, known inhibitors of DENV NS5 would be likely to have efficacy on ZIKV NS5. These comparisons showed a fairly high similarity in shape and structure within inhibitor binding pockets as well as conservation of numerous important residues within those sites. The similarity between the MTase region of ZIKV and DENV NS5 is ~65-70% at aa level and combined with the prediction of similarity presented in the previous study, it has formed the basis of developing ZIKV MTase inhibitors based from the knowledge of DENV-developed inhibitors (174). It has also led to the re-examination of DENV NS5 inhibitors for their efficacy on ZIKV which provided new leads that were more effective in ZIKV than they had been in DENV (175). Potisopon

et al (164) compared the specificity of the RdRp domains of DENV and ZIKV for different modified chain terminators for effective drug discovery. Both polymerases incorporated 2'-C-Me-CMP chain terminators more readily than any other tested modification. The similarities in their interference with the IFN pathway was discussed in 1.6.10.

1.7: Flavivirus treatment and control methods

1.7.1: Drugs

As with many diseases, the symptoms caused by a single flaviviral infection can vary from asymptomatic/relatively mild up to severe symptoms that require hospitalisation. As discussed earlier, DENV infection can range from an entirely asymptomatic infection to a severe haemorrhagic fever with hypovolemic shock. ZIKV can range from asymptomatic infection in the majority of cases to neurological and autoimmune manifestations. Therefore each patient is treated on a case-by-case basis. This is also because there are no specific drugs currently available for flaviviral infections and so treatment of the disease is more a management of the symptoms rather than targeting of the infection itself. In the case of hospitalisation with DENV infection, extensive monitoring is required with fluid replacement with crystalloids or colloids, and potentially serum transfusions if thrombocytopenia occurs (176). The treatments of JEV, YFV, and WNV also involve purely supportive therapies upon hospitalisation. Even ribavirin, the classic wide-spectrum antiviral drug, has been shown to have little effect on flavivirus infections both *in vitro* and *in vivo* and so is not given to patients (177).

There is, therefore, a high demand for flaviviral drugs and a number of potential specific inhibitors have begun to emerge. Non-structural proteins are highly favoured drug targets due to their wide range of activities intracellularly, with emphasis focussed towards those involved in RNA replication, viral capping, and the inhibition of the IFN response, due to their essential nature in the flavivirus lifecycle and previous successes with targeting these processes in other viruses. NS5 and NS3 are particularly favourable targets as core components of the replication complex and so now there are a number of MTase inhibitors (175), polymerase inhibitors (164), chain terminators, and helicase inhibitors in testing (178). Repurposing existing drugs such as Geneticin (179) and Sofosbuvir (180) to treat flaviviral infection is also a possibility. CZS particularly needs specific inhibitors due to the need to treat the infection within a growing foetus. Ribavirin is known to be embryotoxic and teratogenic (181), azauradine is highly toxic to foetuses within model systems (182), and IFN treatments have been flagged as potentially abortifacient (183), meaning that wide-spectrum antivirals cannot be used. The need to prove that the drug has no effects on the foetus' development and health will also add further barriers to make CZS drug development particularly difficult.

1.7.2: Vaccines

With antivirals often being only effective within the first few days of infection, vaccines remain the best route for defence against viruses. There are human vaccines currently available for YFV, JEV, TBEV, and DENV as well as an equine vaccine for WNV (184). The YFV-17D vaccine is the most well-known flavivirus vaccine and was made through serial passage in chicken embryo cells *in vitro* until the virus was attenuated enough to no longer cause human disease. YF-17D produces a very good long-term protective response and has been in use for well over half a century with a very impressive safety record. The YF-17D virus itself is so safe and effective that it formed the basis for the chimeric tetravalent Dengvaxia vaccine virus which came into use in 2015. The threat of ADE has always hindered the development of a DENV vaccine as there needs to be a protective response raised to all four serotypes. The Dengvaxia vaccine, although generating only a mild protective response to DENV-2, was initially deemed to give a good enough response to provide meaningful protection to all four serotypes (12, 13). However, as mentioned in 1.2, since being introduced the vaccine has shown worrying signs of increasing the likelihood of severe disease *via* ADE in countries with low-endemicity as well as giving poor responses in children (14, 15) and has been halted in some countries now over safety concerns. Multiple active vaccines for JEV and TBEV are in clinical use and all require multiple challenges to generate a high seroconversion rate. They are however both effective despite the relatively short lifespan of TBEV protective immunity. WNV vaccines are possible, as shown by the introduction of the equine WNV vaccine, however there is currently no demand from governments to fund and distribute a WNV vaccine as the total numbers of infection and disease severity is considered to be too low.

There are numerous contenders in development as well, including a promising replacement to the Dengvaxia vaccine which produces responses far close to natural infections as well as a stronger DENV-2 response (185). ZIKV has also been very heavily targeted for vaccine production with the recent outbreak (100, 104) revealing the morbidity of CZS and the lack of antiviral drugs available for use in pregnant women so that now many candidates are progressing through early stage trials (186). Despite the lack of government interest in a human WNV vaccine, a number of vaccine candidates have progressed through phase I and phase II trials as WNV cases are increasing and the vector's population has increased in both population and spread in recent years.

1.7.3: Vector control

Flaviviruses are spread *via* vectors, infected blood, and, in the case of ZIKV, sexual contact. The sexual route of ZIKV infection has already been determined to contribute very little to the R_0 value (187) and, although ZIKV did infect the blood supply once during the recent pandemic (31),

there are now layers of protection to stop infected blood from flaviviruses from ever reaching a patient. Therefore the main transmission method for all flaviviruses, and the only one capable of supporting the virus, is vector-borne transmission.

During the 1940's, attempts were made to control mosquito populations in Latin America by using newly developed long-lasting chemical insecticides in focus areas. Whilst this was initially effective, mosquito populations quickly developed resistance to the insecticides and their populations recovered. Insecticides are therefore not a viable option for large scale vector control in modern society due to resistance within mosquito populations, environmental concerns around heavy insecticide use, and the increasing incidence of urban transmission where insecticide use would be heavily restricted (17).

The most promising strategy currently for *Aedes aegypti* control is the introduction of the intracellular symbiont *Wolbachia pipientis* which reduces the lifespan of female mosquitoes by ~50%. As the migration of DENV and ZIKV infection from the midgut of the mosquito to the salivary glands takes ~12 days, the reduction in life expectancy reduces the number of females reaching an age where they can transmit the infection without affecting the mosquito population by killing them before breeding age (188). There is also evidence to show that the presence of *Wolbachia* inhibits the infection of the mosquito and also the transmission of DENV from the salivary glands as well (189). The symbiont is passed maternally and can get into the population if the initial number of introduced infected mosquitoes is high enough by utilising cytoplasmic incompatibility to overcome the fitness cost reduction of being uninfected. This theory was tested and *Wolbachia*-infected *Aedes aegypti* were successfully introduced into the local population within two small areas in Australia (190). *Wolbachia* is a symbiont found in roughly 60% of insect populations and so its use in other species has been questioned, however some research suggests that natural strains of *Wolbachia* are less aggressive than the *w*Mel *Wolbachia* used for these experiments.

Education is also an important factor, especially within urbanised areas where urban-adapted mosquitoes like *Aedes aegypti* breed and thrive. Public education campaigns have been produced to raise awareness of common containers/locations that gather stagnant water and act as a breeding site for these mosquitoes to encourage the public to remove them. Other vector-related theories involve focussing vaccinations on animal hosts to reduce the number of sylvatic-to-anthroponotic infections seen as well as vaccinating those who live/work nearest these animal hosts to also reduce that chance (191).

1.7.4: Host factor targets

Drugs targeting viral proteins give a specific response to a particular infection, however they are expensive to bring to market, rarely work on more than one virus, and there are strong

selective pressures that encourage the development of resistance. Targeting host proteins massively reduces the chance of pathogens developing resistance and targeting host proteins/pathways used by multiple pathogens allows for one drug to inhibit multiple infections. For drugs against host factors to be of interest, the host target must be essential to the virus lifecycle and inhibition of the protein/pathway must not have severe detrimental effects on the patient. Preferably the proteins/pathways to be targeted already are known to be inhibited by modulatory drugs on the market, so the drugs will already be approved as safe for human use and licenced. The concept of host factor drugs has already been proven clinically in HIV patients using CCR5 antagonists (192).

Host factors are key to almost every step of the virus lifecycle and many of these are currently being examined in flaviviruses. DENV cell entry has been shown to be inhibited *in vitro* when exposed to a high level of GAGs, GAG mimics, heparin, or heavily sulphated heparin (193). Flavivirus replication has been shown to be inhibited by; stopping the enzymatic activity of CyPA (172), preventing the binding of elongation factor-1α (EF-1α), polypyrimidine tract-binding protein (PTB), or TIA1 cytotoxic granule associated RNA binding protein/-like 1 (TIA1/TIAR) to flavivirus genomes (178), and through the use of mycophenolic acid to inhibit inosine-5'-monophosphate. Nuclear import inhibition by way of inhibiting importin-α/β, has been shown to inhibit DENV infection (194). Autophagy and the proteasome are important in DENV survival and so inhibitors of these have been shown to severely limit DENV infection *in vitro,* however WNV can be degraded by autophagy and so autophagy inducers limit WNV infection *in vitro* (195).

Host factor targets can also be used to limit specific symptoms from flavivirus infection such as with an *in vitro* experiment showing that the inhibition of metalloproteinases (MMPs) reduces the damage to cell-cell junctions usually seen at the blood brain barrier of patients with neuroinvasive WNV (196). Antagonists to the platelet-activating factor receptor (PAFR) showed in DENV mouse models to reduce thrombocytopenia, vascular permeability, and mortality (197).

1.8: Methods for investigating host-virus protein-protein interactions

The detection of novel protein-protein interactions is an important investigative tool in virus-host studies and there are numerous current methods to do so. All have different pros and cons but none so far are considered proof of interaction and all require further experimentation to prove any potential interactions. Yeast 2 hybrid (Y2H) and yeast 3 hybrid (Y3H) experiments were the original protein-protein interaction techniques which utilise a transactivation protein that has been split and attached to the bait and prey protein so that interaction of the two proteins allows the reformation of the transactivation protein and then transcription of the reporter gene. These techniques, however, are notorious for false positives and require an extensive cDNA library. As the read-out is reporter gene expression and therefore relies on both proteins being localised to the nucleus, false

positives arise from biologically irrelevant interactions between proteins that do not co-localise naturally and also through protein modifications made to localise the protein to the nucleus but that alter the protein's behaviour (198). The split-ubiquitin assay and split-ubiquitin bridge assay was then developed to overcome this problem of forced nuclear localisation, using two fragments of ubiquitin and haemagglutinin (HA) degradation in place of gene transcription. Fluorescence resonance energy transfer (FRET) also uses modified bait and prey proteins to show protein interactions with the donor protein being capable of stimulating the acceptor protein into emitting a particular frequency wavelength if they are within close proximity. This does however require a detailed understanding of the absorbance and emission profiles of both the acceptor and donor and, as with all of these reporter-based bait-prey techniques, it only measures proximity and not direct interaction and so false positives can occur.

Large scale detection techniques that allow for the detection of more than just single protein interactions in each assay are becoming more prevalent. Far western blotting uses a western blot procedure but involves re-naturing the proteins after they have been transferred to a membrane and then incubating the membrane with a bait protein to allow for any complexes between bait and prey to form (199). The bait protein is then probed for during the rest of the western blot procedure to reveal where the bait protein has bound. Mass spectrometry (MS) has become an invaluable tool for protein identification which has allowed for the development of high-throughput techniques though these techniques rely on the correct matching of the peptide fragments to their proteins of origin as well as a comprehensive list of 'possible' proteins in the sample provided by the user for the proteomic programme to compare the spectra to. There are also multiple techniques available for generating the pool of proteins to be used in MS such as co-immunoprecipitation (Co-IP) which uses antibodies against the bait protein bound to beads or within a column which are then used to remove the bait protein from a sample along with any interacting proteins bound to it. The washing steps can remove weak binding partners leading to incomplete interactor lists and at the same time fail to remove strong interactors of the prey proteins themselves or strong non-specific interactors to either protein leading to false positives. Biotin-label identification (BioID) with an APEX (200) or BirA (201) tag allows for permanent labelling of close proximity proteins with molecules of biotin and, as the streptavidin/biotin binding affinity is incredibly high, far stronger detergents can be used compared with standard Co-IP to remove most nonspecific binding to reduce the false positive rate dramatically. As the proteins are labelled permanently, even transient/weak interactors are labelled when they come into proximity with the BirA/APEX tag and will therefore be present in the pulldown sample. Finally, there are *in silico* methods of prediction, which utilise computer modelling techniques to determine likely interactors based on information such as; protein localisation, physicochemical properties, sequence data, structural data, and co-expression data. However these predictions can only be made on proteins that are included in the database used and are well characterised (202).

All except for the *in silico* method requires the experimental expression of the viral protein. Using WT virus for individual protein analysis is not always ideal or possible for some methods and so alternative expression systems of viral proteins are used instead. These expression systems, often of a single viral protein, require some considerations however. Using non-native species for expression can affect post- and co-translational modifications which can affect protein localisation and function. Codon optimisation may be necessary in these cases to account for particular tRNA expression levels but codon sequences can affect co-translational folding as well as translocation of membrane proteins (203). High expression promoters are often used in expression systems however overexpressed proteins will have high levels of nonspecific binding. Some proteins also require host chaperones, the presence of other viral proteins to fold properly, or even specific cleavage to become active. Care must also be taken with the inclusion of tags which can interfere with active regions and affect localisation signals.

1.9: Project aims

The **overall aim** of this project was to generate a list of human proteins interacting with ZIKV NS5 (ZNS5) and compare them to proteins interacting with the DENV-2 and -4 NS5 proteins to identify common proteins and cellular pathways interacting with flavivirus NS5 proteins.

The **specific aims** were to:

1) use ZNS5-FLAG and DENV NS5-FLAG expressing cell lines; both already available, for comparative pull-down and tandem mass tagging/mass spectrometry (TMT-MS) analyses.
2) produce a ZNS5-BirA expressing cell line and devise and implement a method for whole cell streptavidin-bead pulldowns such that the recovered proteins can be analysed by TMT-MS.
3) undertake a comparative analyses of the proteomic datasets produced in Aims 1 and 2 using commercially available proteomic programmes, and online bioinformatic resources, to identify proteins commonly and distinctly interacting with ZNS5 and DENV NS5 as well as those proteins interacting with ZNS5-FLAG and ZNS5-BirA. Over the course of this investigation, proteomic datasets generated using different bioinformatic analyses software (Proteome Discoverer (PD) and MaxQuant) will also be compared to identify the differences in MS proteomic programmes and highlight the inconsistencies and flaws within the current high-throughput proteomic technology.

2: Methods and materials

2.1: Cell culturing

2.1.1: Cell culture conditions

Cells were incubated at 37^0C and 5% CO_2 in sterile T25/75/175/225 flasks that contained a growth media (GM) which consisted of: Dulbecco's Modified Eagle's Medium (DMEM) + Glutamax (Lonza, Basel, Switzerland), 10% foetal bovine serum (FBS) (Life Technologies, Paisley, UK), and 0.1mM non-essential amino acids (Life Technologies). Cells were washed with phosphate-buffered saline (PBS) (Lonza) and then detached using 10% trypsin when they reached high confluency before being passaged to a new flask. Stably transfected cell lines also received 150µg/ml hygromycin/3.5µg/ml puromycin and 15µg/ml blasticidin (Sigma Aldrich, Gillingham, UK) added once every two passages and every four passages respectively for selection purposes. All work with live cells took place in a sterile laminar flow hood.

2.1.2: Cell lines

Table 1: *Cell line details:*

Cell line name	**Cell line description**	**Production credit**
293-Flp	HEK 293 Flp-In™T-REx™ cells	-Life Technologies
D2NS5-FLAG	HEK 293 Flp-In™T-REx™ cells expressing DENV2 NS5-FLAG	-Life Technologies -Dr Holga Hannemann (Dr Andrew Davidson's lab)
D4NS5-FLAG	HEK 293 Flp-In™T-REx™ cells expressing DENV4 NS5-FLAG	-Life Technologies - Dr Holga Hannemann (Dr Andrew Davidson's lab)
ZNS5-FLAG	HEK 293 Flp-In™T-REx™ cells expressing ZIKV NS5-FLAG	-Life Technologies -Mr Josh Lee (Dr Andrew Davidson's lab)
ZNS5-BirA	HEK 293 Flp-In™T-REx™ cells expressing ZIKV NS5-BirA*-containing HEK 293 Flp-In™	-Life Technologies -Mr Jack Hales/Dr Lisa Stevens (Dr Andrew Davidson's lab)

2.2: DNA/RNA techniques

2.2.1: DNA gel electrophoresis

1% agarose gels (w/v) (0.5µg/ml ethidium bromide) were made by dissolving 1g agarose powder (Promega, Wisconsin, USA) in 100ml 1xTBE buffer (Appendix A) under heat and then mixing in 50µg ethidium bromide (Invitrogen, California, USA) by gentle swirling. It was then poured into a gel mould and left to set. Once set, the gel was placed in a Sub Cell GT Cell horizontal electrophoresis tank (Bio-Rad, California, USA) and submerged in 1xTBE buffer (0.00005% 10mg/ml ethidium bromide (v/v)). DNA samples were mixed with 6xDNA loading dye (Thermo Fisher Scientific, Massachusetts, USA) at a 5:1 ratio and were then loaded into individual wells alongside GeneRuler 1kb Plus DNA ladder (Thermo Fisher Scientific) before being run at 100V for ~30-60min. The final gel was then imaged using a UVP Biodoc-IT™ System Ultraviolet Transilluminator UV spectrometer (UVP, California, USA).

2.2.2: DNA concentration determination

The DNA concentration of samples was determined using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific). Small aliquots of the samples were then run on an electrophoresis gel alongside 5µl GeneRuler 1kb Plus DNA ladder (Thermo Fisher Scientific) so that the purity of the sample could be assessed and an estimation of the DNA concentration of the desired band could be determined. It was then compared back to the NanoDrop™ results.

2.2.3: Primer design and synthesis

pcZNS5-BirA was sequenced using five forward oligonucleotide primers post-midi-prep to ensure that the cloning procedure had not led to mutation of the ZNS5-BirA sequence. Primers were designed every ~700 nucleotides to cover the entire sequence using the same gene sequence used to design the codon-optimised ZNS5 and BirA genes synthesised by Invitrogen. The primers were synthesised by Eurofins Genomics (Luxembourg City, Luxembourg).

2.2.4: Primer sequences

Table 2: *Primer details:*

Primer name	Primer sequence (5′-3′ direction)	Target gene/gene fragment	Forward/Reverse
prStart	CGACTCACTATAGGGCGAATTCGG	ZNS5	Forward
pr582	GATGGAAACCCTGGAACGGC	ZNS5	Forward
pr1260	GAAAGAGTGGAAAACCGCCG	ZNS5	Forward
prMid	CAAGTACACCTACCAGAACAAGG	ZNS5	Forward

pr1980	GAAGAGAATGGCCGTGTCCG	ZNS5	Forward
pr2647	GAGAAGTACATGGACTACC	ZNS5	Forward
prB621	CATGGCCATGAGAAGAGTGG	BirA*	Forward
prEnd	GTCACTTCTCTGCGCTTCTCAGGG	ZNS5	Reverse
prMidRev	GCTGCACCACCAGGTTGGTGAAGG	ZNS5	Reverse

2.2.5: DNA sequencing

DNA sequencing was done externally by Eurofins Sequencing. 15µl 10ng/µl sample DNA was sent off accompanied by 15µl 10pmol/µl primer DNA per primer per reaction. The results of the sequencing were BLAST searched against the original sequences which showed no mutations. The sequence of the primers are provided above in table 2 in 2.2.4 and the placement of all primers within the gene as well as the sequencing results are provided in appendix B.

2.2.6: RT-PCR

RNA used in the RT-PCR procedure was isolated from samples (see 2.3.1) using the SV Total RNA Isolation kit (Promega) following the protocol provided. The primers were designed as described in 2.2.3 and are shown in table 2 in 2.2.4.A one-step, two fragment RT-PCR was performed using the One *Taq*® RT-PCR kit (New England Biolabs, Massachusetts, USA) according to their protocol by creating a mastermix consisting of: ~1µg RNA, 2µl 25xOne*Taq* Enzyme Mix, 25µl 2xOne*Taq* Reaction Mix, 2µl 10µM forward primer, 2µl 10µM reverse primer and made up to 50µl with ddH_2O. A GS1 thermocycler (G-Storm, Somerset, UK) was then used to perform the run cycle shown in Figure 10. After DNA electrophoresis confirmed that the two step RT-PCR was successful, the samples were purified using the reagents and protocol from the GeneJET PCR Purification kit (Fermentas, Leicestershire, UK).

Figure 13:

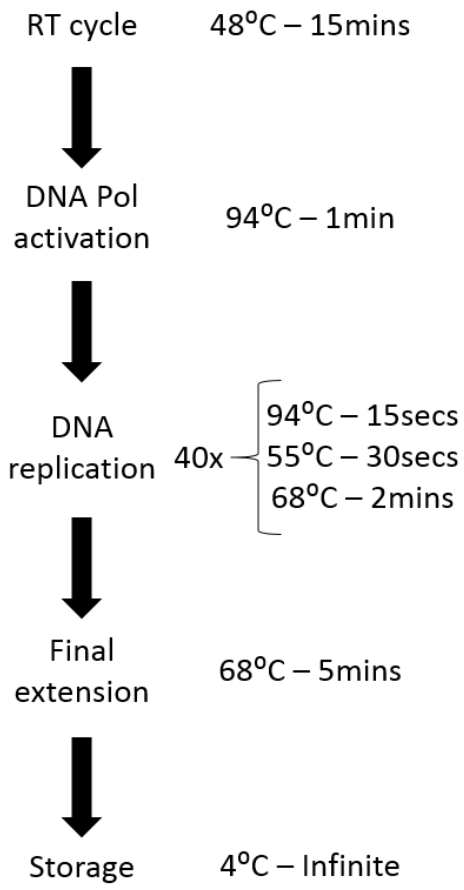

Figure 13: *Thermocycler programme for two-fragment, one step RT-PCR using One Taq RT-PCR kit (New England Biolabs)* – The thermocycler sequence was programmed in accordance with the guidelines for the One Taq RT-PCR kit protocol. As both fragments were ~2kb in size, 2 minutes was selected as the extension time (1 min per kb) and 55°C was chosen as the annealing temperature based on the primer characteristics

2.2.7: Transfection of 293-Flp cells (transient)

The pGK-Helper plasmid (ThermoFisher Scientific), and pcZNS5-BirA were mixed at the ratios of 9:1 and 11.5:1 (µg/µg) and then added to an appropriate amount of TurboFect (ThermoFisher Scientific, following the standard recommended transfection procedure) and vortexed before being added dropwise onto a culture of ~70-80% confluent 293-Flp cells in a 24-well plate. After overnight incubation, doxycycline was then added to the media to a final concentration of 1µg/µl and the cells incubated for 4d for adequate expression of ZNS5-BirA.

2.2.8: Transfection of 293-Flp cells (stable)

The transfection was as described in 2.2.7 but after overnight incubation with the plasmids and Turbofect. The 293-Flp cells were then detached and plated into 10cm dishes. After overnight incubation, hygromycin was added to the media to a final concentration of 150ug/ml and hygromycin resistant cell colonies selected for over the course of 3-4 weeks. Hygromycin resistant

cell colonies were then removed from the 10cm dish and grown up sequentially in a 24-well plate, T25 flask, and then T75 flask for culturing and confirmatory tests.

2.2.10: Bacterial transformation

50µl α-Select Silver Efficiency *E.coli* (Bioline, Toronto, Canada) were incubated on ice for 30min with 1-10ng of plasmid DNA. The bacteria were then heat shocked at 42ºC for 30-45sec and then immediately placed on ice for 2min. 1ml of warm Luria broth (LB) (Appendix A) was added before being incubated for 1h at 37ºC with shaking at 225rpm. The bacteria were then plated onto warmed LB agar (Appendix A) containing 50µg/ml zeocin (w/v) (Thermo Fisher Scientific) and then incubated overnight at 37ºC. Resulting colonies were then picked and grown in LB.

2.2.11: Growth/selection of bacteria

Transformed *E.coli* were plated on LB agar (Appendix A) under the selection pressure of 50µg/ml zeocin (w/v) (Thermo Fisher Scientific) and incubated overnight at 37ºC. A colony was then picked and grown in LB (Appendix A) with 50µg/ml zeocin (w/v) (Thermo Fisher Scientific) at 37ºC with shaking at 225rpm.

2.2.12: Restriction enzyme (RE) digest

RE digests were performed using HindIII and BamHI kits (New England Biolabs) according to the protocols provided. 1µg of sample DNA was mixed with 1µl HindIII (~10 units), 1µl BamHI (~10 units), and 5µl 10x NEbuffer 2.1 (New England Biolabs) and made up to 50µl with ddH_2O. The sample was then incubated for 1h at 37ºC. The REs were removed using a GeneJET PCR purification kit (Fermentas).

2.2.13: Ligation

The ligation was done in accordance with the T4 DNA Ligase protocol provided with the T4 DNA Ligase kit (New England Biolabs). A reaction mix consisting of: 25µg of vector DNA, 75µg of insert DNA, 2µl 10xT4 DNA Ligase buffer (New England Biolabs), and 1µl T4 DNA Ligase was made with ddH_2O to make the reaction volume to 20µl. The reaction mix was incubated at room temperature for 10min and then heat inactivated at 65ºC for a further 10min.

2.3: Protein techniques

2.3.1: Cell lysis

The cells were detached using 10% trypsin (in PBS), centrifuged at 150 x *g* for 5min at room temperature, washed with PBS, and then centrifuged at 150 x *g* for 5min a second time. The PBS was removed and 0.5ml of 2x sample buffer (Appendix A) was added to each pellet and mixed

using a pipette. A 21 gauge needle and a 1ml syringe were then used to passage the cell lysate 5 times up and down. The resulting sample was heated at 95^0C for 5min and then stored at -20^0C.

2.3.2: BCA assay

The BCA assays were done using the Pierce Microplate BCA Protein Assay Kit (Reducing Agent Compatible) (ThermoFisher Scientific) according to the protocol provided in the kit and using the cell lysate attained using the method described in 2.3.1.

2.3.3: Immunofluorescence assay (IFA)

Sterile glass coverslips were placed in 24 well plates, washed with PBS (Lonza), incubated with poly-D-Lysine (0.1mg/ml in PBS) for 5min, and then washed twice with PBS. Cells were grown in GM + doxycycline (1µg/ml) for 3d before being seeded at roughly $5x10^4$ cells per well. These were incubated in GM + doxycycline (1µg/ml) overnight before fixation in either ice cold methanol for 5min, or 4% paraformaldehyde for 5min followed by permeabilization with 1% (w/v) Triton-X 100 for 5min. The fixed cells were blocked in 1% (v/v) FBS (in PBS) on a plate rocker for 1h at room temperature. The blocking solution was removed and the coverslips coated in 100µl of primary antibody solution at the recommended IFA concentration (in 1% FBS (in PBS), (antibody information in 2.10) for 1h at room temperature. The primary antibody was removed and the coverslips were washed 4 times in PBS for 5min per wash before 100µl of secondary antibody solution was added at the recommended concentration (in 1% FBS (in PBS)). After removal of the secondary antibody the coverslips were washed 4 times in PBS for 5min per wash and then mounted onto glass slides using VectaShield® mounting media (Vector laboratories, California, USA) and nail varnish to seal the glass coverslips. Slides were then imaged under a DM IRB inverted epifluorescence microscope (Leica, Wetzlar, Germany) at 60x or 100x magnification using the Leica Application Suite Advanced Fluorescence programme to capture and analyse the images.

2.3.4: Coomassie blue staining

Samples were prepared by adding appropriate quantities of cell lysate (as determined by BCA assay) to 5% β-mercaptoethanol (B-MERC) (made up in 2x sample buffer containing bromophenol blue). The samples were then heated at 95^0C for 5min before being loaded into a reducing and denaturing 10% SDS-Page gel with 5% stacking gel and run at 100V for ~1h. The gel was then transferred to a container containing Coomassie Brilliant Blue dye (ThermoFisher Scientific) (Appendix A) and placed on a plate rocker for 1h. It was then placed in Coomassie destain solution (Appendix A) overnight.

2.3.5: HRP-conjugated western blot

The samples were set up and run as in 2.3.4 but instead of being placed in Coomassie brilliant blue (ThermoFisher Scientific) after running, the gel was placed in 1x transfer buffer (Appendix A) for 5min and then transferred to an Amersham™ Hybond™ P0.45 PVDF nitrocellulose blotting membrane (GE Healthcare, Buckinghamshire, UK) using a Trans-Blot® Semi-Dry Transfer Cell (Biorad, California, USA) at 15V for 1h between two pieces of blotting paper that had been dampened in 1x transfer buffer. This type of transfer was done when examining proteins that ranged ~25-150kDa in size. Those larger than 150kDa were transferred using a wet transfer blot technique as described in 2.3.7. Before transfer, the nitrocellulose membrane was washed in 100% methanol for 1min, distilled water (ddH_2O) for 2min, and 1x transfer buffer for 5min. After transfer the membrane was placed in a 50ml Falcon tube containing 5ml of blocking solution (Appendix A) and rotated on a tube roller for 1h at room temperature. The membrane was then cut to detect proteins with different molecular weights as needed. Then each filter fragment was placed in a 50ml falcon tube containing an appropriate concentration of primary antibody diluted in 3 ml of blocking solution and placed on a tube roller for 1h at room temperature. The membranes were washed in PBS containing 0.1% (v/v) Tween, 4 times for 5min each wash before being placed in 50ml falcon tubes containing an appropriate concentration of HRP-conjugated secondary antibody diluted in 3 ml of blocking solution and placed on a tube roller for 1h at room temperature. The membranes were washed again in PBS containing 0.1% (v/v) Tween, 4 times for 5min each wash before being coated in ECL reagent (Insight Biotechnology, Wembley, UK) for 1min. The excess fluid was removed and the membranes were placed inside a developing cassette and exposed to Hyperfilm ECL (Amersham, GE Healthcare) once in a dark room. The films were then developed in a Konica SRx-101A film processor.

2.3.6: Fluorescent western blot

The fluorescent Western Blot was run in the same way as described in 2.3.5 except the Amersham™ Hybond™ P 0.45 PVDF nitrocellulose blotting membrane (GE Healthcare) was replaced with an Immobilon®-Fl Transfer Membrane (Li-Cor, Nebraska, USA). The secondary antibodies are attached to fluorescent molecules instead of being conjugated to HRP, causing them to fluoresce under particular wavelengths of light depending on the antibodies used. They should also be made up in blocking solution (Appendix A) containing 0.01% SDS. Alternatively, a fluorescently-labelled binding partner can be used in place of an antibody such as the Streptavidin Alexa Fluor™ 488 (Thermo Fisher Scientific) which is the binding partner for biotin.

2.3.7: Wet transfer

The protocol used for wet transfer allowed for the transfer of proteins up to ~500kDa within 1h. The samples were incubated at 95ºC for 10min with 2x sample buffer (5% β-mercaptoethanol) and

then run on NuPAGE™ 3-8% Tris-Acetate gels (Thermo Fisher Scientific) with a high molecular weight marker; HiMark™ prestained protein standards (Thermo Fisher Scientific). The gel was run in a Mini Gel Tank (Thermo Fisher Scientific) in 400ml 1x NuPAGE™ Tris-Acetate SDS Running Buffer at 150V for 1h. The Amersham™ Hybond™ P0.45 PVDF nitrocellulose blotting membrane was equilibrated by soaking it in 100% methanol for 1min, ddH_2O for 2min, and 1x NuPAGE™ Transfer Buffer (Thermo Fisher Scientific) for 5min. The gel was soaked in 1x NuPAGE™ Transfer Buffer (Thermo Fisher Scientific) for 15min which was made up from 20x NuPAGE™ Transfer Buffer (Thermo Fisher Scientific), ddH_2O, and 10% methanol/ 20% methanol for one gel being transferred or two respectively. The transfer 'sandwich' was then made up according to Figure 14 and placed in the Mini Trans-Blot® Electrophoretic Transfer Cell (Biorad) with 1x NuPAGE™ Transfer Buffer (Thermo Fisher Scientific) (+ an appropriate concentration of methanol) filling the transfer tank. A small magnetic stirrer at the bottom of the tank and an ice pack in the transfer buffer is essential to keep the ion concentration even and the temperature of the transfer components low. The transfer settings are described in table 3. Once the transfer was complete, the Western blot procedure continued post-transfer as described in 2.3.5.

Figure 14:

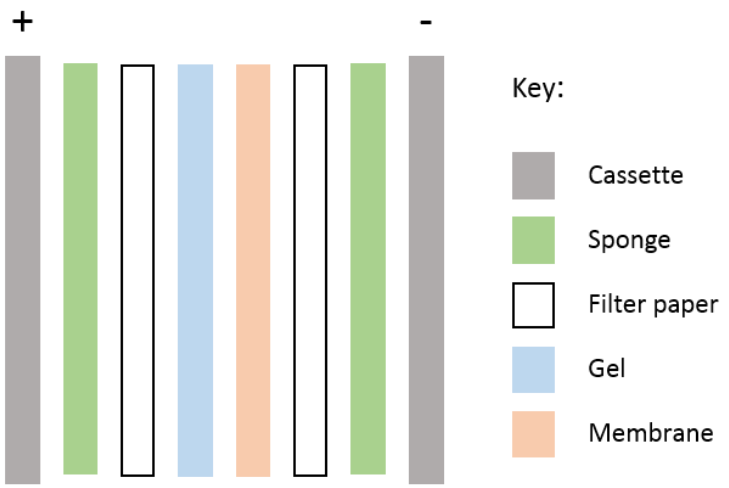

Figure 14: *Wet transfer 'sandwich'* – All components of the wet transfer 'sandwich' must be already wet with transfer buffer before construction of the 'sandwich' and the gel must be placed towards the cathode (+) to allow for migration of proteins into the membrane in the direction of the anode (-) with a piece of blotting filter paper on either side as well as a wet transfer blotting sponge

Table 3: *Wet transfer settings:*

Number of gels	Transfer time	Voltage (V)	MilliAmperes (mA)
1	1h	100	350
1	Overnight (~16h)	30	90
2	1	100	375-450
2	Overnight (~16h)	30	100

2.3.8: FLAG-bead pulldown

Cells were grown to around 70% confluency in T225's and doxycycline added to the media to a final concentration of 1ug/ml for 4d before cell harvesting. The cells were detached by gentle scraping and then harvested by centrifugation at 300 x *g* for 5min at 4ºC. The cells were then washed twice with PBS with centrifugation at 300 x *g* for 5min at 4ºC each time. The cell pellets were then resuspended in 1ml ice cold lysis buffer (Appendix A) per T225 used and left on ice for 30min with mixing by pipetting every 10min. The samples were then sonicated 4x (pulse 5 secs on, 10 sec pause, Amp 50%) with a Q125 Sonicator (QSonica, Connecticut, USA), centrifuged at 14,000 x *g* for 15min at 4ºC, and then added to equilibrated Pierce Control Agarose Resin (Thermo Fisher Scientific) to be mixed on an end-over-end rotor for 1h at room temperature. The samples were then centrifuged at 3000 x *g* for 1min and the supernatant of each was added to FLAG magnetic beads that had been equilibrated by washing twice in 10x resin volume equilibration buffer (Appendix A) with 50µl of FLAG beads per ml of lysate used. The samples were then placed on an end-over-end rotor overnight at 4ºC. Using a magnetic Eppendorf tube holder, the supernatant was removed and washed 3 times in 5x bead volume Wash Buffer (Appendix A) before resuspending the FLAG beads in 20µl of Wash Buffer and storing in -70ºC until being sent for TMT labelling. The samples and all reagents were kept on ice throughout except where specified.

2.3.9: Biotin labelling

To induce biotin labelling, cells expressing BirA or BirA fusion proteins were incubated at 37ºC and 5% CO_2 with 2µg/ml biotin for 14-18h.

2.3.10: Streptavidin pulldown

The protocol for the streptavidin bead pulldown was adapted from protocols used in previous BioID pulldown studies to suit large scale whole cell lysates (204, 205). The steps are similar to the FLAG pulldown protocol in 2.7 but, due to the strength of the streptavidin-biotin binding, the washing buffers were much more stringent than those used in the FLAG pulldown (the recipes of which can be found in the appendix). The cell lysates were prepared in the same way as for the FLAG protocol except the cell lysates (derived from 3x 80% confluent T225's) were incubated with rotation overnight at 4ºC with 50µl Streptavidin Sepharose High Performance beads (GE Healthcare). After overnight incubation with the sample, the beads were washed for 10min with 1.5ml of Wash Buffer 1, 2, 3, then 4 (Appendix A) with 1min of centrifugation at 1000 x *g* between each wash to collect the beads and allow for the removal of the previous wash buffer. The final beads were then resuspended in 20µl of Wash Buffer 4 and then stored at -70ºC until analysis.

2.3.11: Protein G-bead pulldown

The cell lysates were prepared using the same protocol as described for both the FLAG and Streptavidin pulldowns. The lysates were then incubated with rotation overnight at 4ºC with the selected antibody at the concentration specified by the antibody manufacturer. After the overnight incubation, the samples were incubated with rotation for 1h at room temperature with equilibrated Pierce™ Protein G Magnetic Beads (Thermo Fisher Scientific) at 50µl per ml of cell lysate. Using a magnetic Eppendorf tube holder, the supernatant was removed and the beads were washed 3 times in 5x bead volume Wash Buffer (Appendix A) before resuspending the beads in 20µl of Wash Buffer and storing in -70ºC

2.3.12: Antibody and binding partner information

Table 4: *Antibody details:*

Antibody/binding partner target	Company	Identifier/ catalogue number	Species	Experiments used in	Concentration used
FLAG epitope	Sigma Aldrich	F1804	Mouse	IFA, WB	IFA: 1:500-2000 WB: 1:1000
GAPDH	Sigma Aldrich	G9545	Rabbit	WB	WB: 1:5000
ERC1	Abcam	ab50312	Mouse	WB	WB: 1:2000
Biotin (Streptavidin Alexa Fluor™ 488)	Thermo Fisher Scientific	S11223	-	WB	WB: 1:5000
ZNS5	GeneTex	GTX133312	Rabbit	IFA, WB	IFA: 1:500 WB: 1:100-10,000
BirA	Aviva Systems Biology	OARA01977	Rabbit	IFA, WB	IFA: 1:1000 WB: 1:5000
UBR5	Abcam	ab4376	Goat	WB, IP	WB: 1:500 IP: I:200
Cdk1	Call Signalling Technology	#9116	Mouse	WB, IP	WB: 1:1000 IP: 1:100
STAT2 (non-phosphorylated) B3	Santa Cruz Biotechnology	sc514193	Mouse	WB	WB: 1:100

anti-Rabbit-Ab (Alexa Fluor™ 568)	Abcam	ab175471	Goat	IFA	IFA: 1:2000
anti-Mouse-Ab (Alexa Fluor™ 568)	Life Technologies	A11031	Goat	IFA	IFA: 1:1000
Anti-Rabbit-Ab (HRP conjugated)	Santa Cruz Biotechnology	sc-2054	Goat	WB	WB: 1:5000
anti-Mouse-Ab (HRP conjugated)	Santa Cruz Biotechnology	sc-2055	Goat	WB	WB: 1:5000
anti-Goat-Ab (HRP conjugated)	Santa Cruz Biotechnology	sc-2056	Donkey	WB	WB: 1:5000
anti-Mouse-Ab (Alexa Fluor™ 800CW)	Li-Cor	926-32210	Goat	WB	WB: 1:10,000
anti-Rabbit-Ab (Alexa Fluor™ 800CW)	Li-Cor	926-32211	Goat	WB	WB: 1:10,000
anti-Mouse-Ab (Alexa Fluor™ 680LT)	Li-Cor	926-68020	Goat	WB	WB: 1:10,000
anti-Rabbit-Ab (Alexa Fluor™ 680LT)	Li-Cor	926-68021	Goat	WB	WB: 1:10,000

2.4: Proteomics and bioinformatic techniques

2.4.1: TMT-labelling and LC-MS/MS

100µg of each sample was digested with 2.5µg of trypsin overnight at 37ºC and labelled with TMT reagents. The samples were pooled and evaporated until dry before being resuspended in buffer A (Appendix A) and then fractionated in an Ultimate 3000 liquid chromatography system with an XBridge BEH C18 column (Waters, Herts, UK). Proteins were eluted with an increasing gradient of buffer B (Appendix A) from 0-95% for 60min. These collected fractions were then evaporated until dry again and resuspended in 1% formic acid before being analysed in an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) using the Xcalibur 2.0 software (Thermo Fisher Scientific). The peptides were detected in an FT-IT MS/MS manner with MS1 being performed at a resolution of 120,000, a max injection time of 50ms, an automatic gain control of

200,000, and a normalised collision energy of 55%. The monoisotopic precursor selection setting was used to select the peaks that proceeded to MS2 which were detected with a quadrupole mass filter ion trap at a width of 1.2m/z. The samples for MS2 were processed before detection with a collision-induced dissociation collision energy of 35%, a max injection time of 70ms, and an automatic gain control target of 5000.

2.4.2: Bioinformatic analysis

The TMT-LC-MS/MS data was initially processed through a quantitative proteomic software package, either PD or MaxQuant, to identify the proteins of origin for each peptide fragment and determine their relative quantities. This data was then further processed in Perseus, a modified SPSS programme designed for proteomic data. The statistically analysed data was then exported to Microsoft Office Excel where it could be processed by hand to create spreadsheets of high confidence hits based on an increase of at least 1.5 fold compared with control samples as shown by t-test previously performed in Perseus and those with statistical significance based on the p value of that t-test. Some consideration was taken for proteins slightly under this cut-off point if there were mitigating circumstances. High confidence lists for the potential protein interactors were compared in Microsoft Office Excel and were also analysed individually using the online bioinformatic resources STRING and DAVID that were highlighted in the literature (206) as being effective in pathway, Gene Ontology, and protein cluster analysis, to enable us to compare not only the individual proteins but also the processes and pathways that were targeted by each viral NS5 protein.

3: Results

3.1 Experimental background to the study

Prior to the commencement of this study, experiments had been done in the laboratory to examine the human cellular protein interactome of the DENV-2 and -4 NS5 proteins (D2NS5 and D4NS5 respectively). For these experiments 293-Flp cell lines had been constructed that stably expressed the D2/4NS5 genes under control of an inducible promoter (159). To regulate the expression of the NS5 genes, a $TetO_2$ operator was used, which allowed inducible transcription in response to the addition of tetracycline/doxycycline to the cell culture media. Compared to transient expression systems, the use of the 293-Flp cells facilitated expression of the NS5 genes at more physiological levels in the entire cell population. The NS5 genes were expressed in isolation from the other viral genes in order to identify protein interactions that were mediated solely by the NS5 proteins and not other viral proteins that may interact with NS5. To allow efficient purification of the NS5 proteins and their cellular interaction partners, sequence encoding a FLAG epitope tag was fused to the 3′ end of the NS5 gene sequences.

It was decided to adopt a similar strategy for expression of the ZNS5 gene for interactome analysis. The production of a 293-Flp cell line that expressed the ZNS5 gene fused to a sequence encoding a FLAG tagged epitope would allow a direct comparison with the cellular interactomes of the DENV NS5 proteins which were also FLAG epitope tagged and expressed in 293-Flp cells. The FLAG epitope tag was used, once again, so that experiments could be done under the same conditions using the same reagents/antibodies for protein purification. A human-codon-optimized ZNS5 gene fused to a 3′ terminal FLAG tag sequence (designed by Dr A. Davidson; sequence in Appendix B) was synthetically constructed from oligonucleotides and then ligated into the plasmid pMK-RQ(kanR) (Thermo Fisher Scientific). The DNA was sequenced and determined to be accurate before being delivered. In addition to the use of the FLAG tag for protein purification, a BirA* tag was also selected for its functionality as a proximity labelling tag, which would potentially allow the detection of proteins that interacted even transiently/weakly with ZNS5. These interaction partners could then be compared to those using the traditional FLAG epitope tag pulldowns. The plasmid pcDNA5FrtTolntr-S497, which contained a BirA*-tagged insert (the insert corresponded to the E gene of Middle East respiratory coronavirus (MERS-CoV) (Thermo Fisher Scientific) was also purchased. Using a HindIII/BamHI digest, the ZNS5 gene was then removed from pMK-RQ(kanR) and ligated into the corresponding sites of pcDNA5FrtTolntr-S497, such that the ZNS5 sequence was fused at the 3′ end to a sequence encoding the BirA* tag rather than a FLAG tag. A vector map for both plasmids is shown in Figure 15B. The design of the plasmids, the generation of a ZNS5-FLAG stable cell line and the preparation of the ZNS5-BirA plasmid were all completed before the start of this project (done by Dr A. Davidson / Mr Joshua Lee, Research Technician in the laboratory).

3.2: Introduction to the 293 Flp-In™ system

The 293-Flp-in™ system is a commercially available system to allow for efficient production of cells stably expressing exogenous genes. As shown in Figure 15, the main constituents of the 293-Flp-in™ system are the FRT (Flp recombination target) site and the zeocin resistance gene which allows for selection of cells with an intact Flp-in site. The FRT Flp-in plasmid inserts *via* the action of the Flp recombinase which is supplied *in trans* by a helper plasmid expressing the Flp recombinase enzyme. This cleanly inserts all of the plasmid DNA into the chromosomal DNA of the 293-Flp cells as shown below. The FRT Flip-in plasmid can be designed to contain a particular gene of interest or one can be added using restriction enzyme digests after ordering the plasmid backbone. The gene of interest itself is expressed using a high expression promoter; P_{CMV}, and this is under the control of the tetracycline/doxycycline operator $TetO_2$. Once inserted into the chromosome of the 293-Flp cells, the DNA is stable as long as hygromycin selection occurs once every 2-4 passages.

Figure 15:

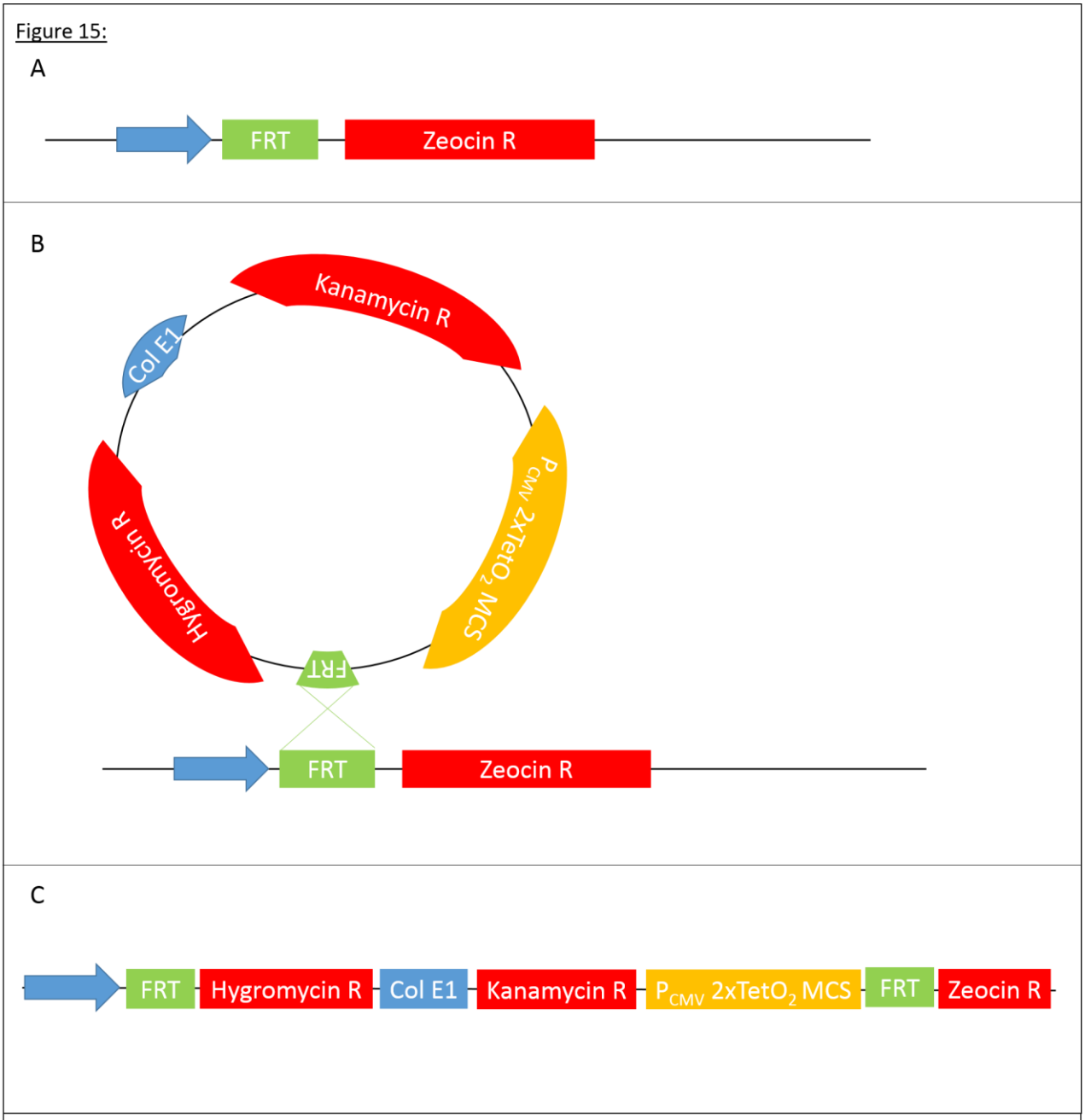

Figure 15: *The 293 Flp-In™ system - A*: the basic components of the ThermoFisher Scientific 293Flp-In™ recombinase site found in the genome of the 293Flp-In cells. *B*: the recombinase event with a consensus FRT Flip-In plasmid (not shown: the Flp recombinase enzyme supplied *in trans* by a helper Flp recombinase plasmid.) *C:* After the recombination event, the plasmid inserts into the chromosomal DNA and remains stable as long as selection pressures are still applied. Expression of the chosen gene of interest takes place with a high expression promoter under the control of a tet-on system
FRT – Flp recombinase target, ColE1 – Col E1 Ori, P_{CMV} – CMV promotor, 2xTetO$_2$ – Tetracycline/Doxycycline-dependent promotor, MCS – Multiple cloning site

3.3: Production and testing of a ZNS5-FLAG cell line

Prior to the commencement of the project a stable ZNS5-FLAG expressing cell line was produced (by Mr Joshua Lee). Briefly, the ZNS5-FLAG FRT Flp-in plasmid described above was used to transfect 293-Flp cells according to the protocol in 2.2.8 using ratios (µg:µg) of 11.5:1 and 9:1 of helper plasmid:ZNS5-FLAG FRT Flp-in plasmid. Approximately 2d after transfection, the

transfected 293-Flp cells were then cultured using media containing 150ug/ml hygromycin to select for stable recombinants. Colonies were then isolated and cultured and then tested for production of the ZNS5-FLAG protein prior to this project.

In order to determine the timing of induction for optimal protein production, ZNS5-FLAG expressing cells were grown in media containing 1µg/ml doxycycline for various times and then examined by IFA and western blot analysis. Every 2 days for 8 days the cells were either fixed or harvested and used to make total cell lysates for IFA and Western blot analysis respectively. The IFA samples were kept in PBS at 4 ºC after 4% paraformaldehyde fixation and the cell lysates were stored at -20 ºC until all samples were collected. The IFA was performed (according to 2.3.3) using an anti-FLAG antibody and analysed by widefield fluorescence microscopy. Phase contrast images were also taken as it was observed that cell health declined after 4d of doxycycline exposure. The results (Figure 16) showed that very minimal induction of ZNS5-FLAG was detected after 2 days whereas day 4 to day 8 showed far more detectable levels of ZNS5-FLAG. There was however a noticeable decrease in cell health after day 4 with day 6 and day 8 showing a higher number of rounded cells.

Figure 16:

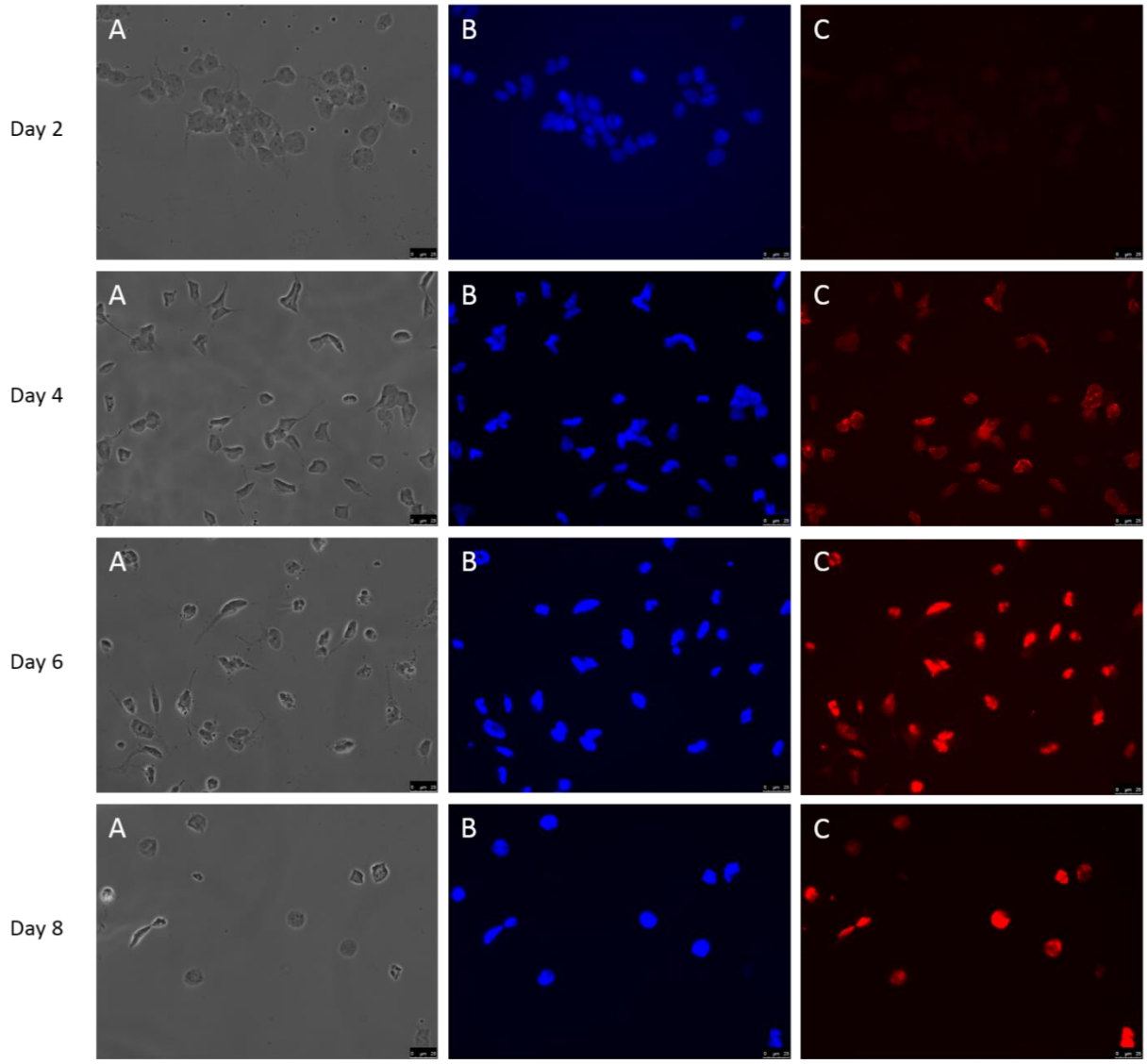

Figure 16: *Time lapse IFA of ZNS5-FLAG cell lines exposed to doxycycline and probed with FLAG antibody* – ZNS5-FLAG cells were incubated with doxycycline (1µg/ml) for 8 days with samples fixed after 2 days, 4 days, 6 days, and 8 days which were then used in immunofluorescent staining with VectaShield® mounting media containing DAPI blue. The resulting images were taken on a 60x widefield lens in; *A*) phase contrast, *B*) under blue fluorescence for DAPI nuclear staining, and *C*) observing red fluorescence for anti-mouse Ab (568) binding to anti-FLAG Ab.

The amount of protein in the cell lysates were then quantitated using a BCA assay and then equal quantities of protein were analysed by western blot using an anti-FLAG antibody. A band was detected at ~105kDa, the predicted size of ZNS5-FLAG, in all cells incubated with doxycycline (Figure 17). The blot was also probed with an anti-GAPDH antibody. A band was detected at ~37 kDa; the expected size of GAPDH. It appeared from the western blot that the peak time of ZNS5-FLAG accumulation was 6d of induction with doxycycline.

Figure 17:

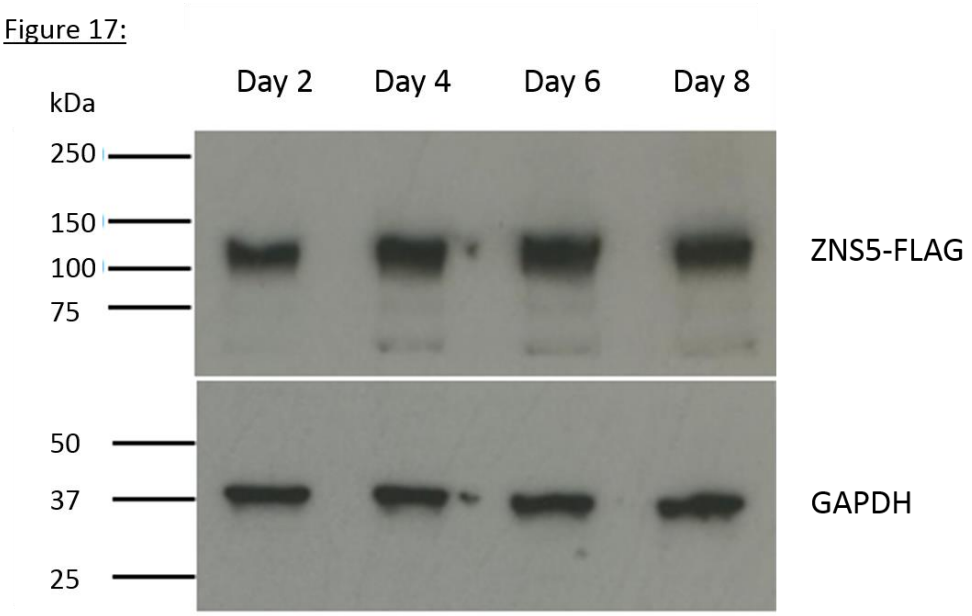

Figure 17 : *Time lapse western blot of ZNS5-FLAG cell lines exposed to doxycycline and probed with FLAG antibody* - ZNS5-FLAG cells were incubated with doxycycline (1µg/ml) for 8 days with samples lysed after 2 days, 4 days, 6 days, and 8 days which were then used in HRP-conjugated western blotting using anti-FLAG and anti-GAPDH antibody. The film was developed in a Konica film developer.

It was decided that ZNS5-FLAG was being expressed as desired and that 4d of doxycycline exposure (1µg/ml) were optimal for cell health and provided ample protein expression. The localisation of ZNS5 was percieved to be highly nuclear in Figure 18 which fit with previous descriptions of ZNS5 localisation. To ensure that the ZNS5-FLAG gene was being expressed solely under the induction of doxycycline and to test the specificity of the anti-FLAG antibodies, another IFA was performed comparing ZNS5-expressing cells with no doxycycline exposure and those exposed to doxycycline for 4d. 293-Flp cells were also treated in the same way as a control group for the experiment. As shown in Figure 18, the red fluorescence indicative of the presence of the FLAG epitope was only present in doxycycline-induced ZNS5-FLAG cells.

Figure 18:

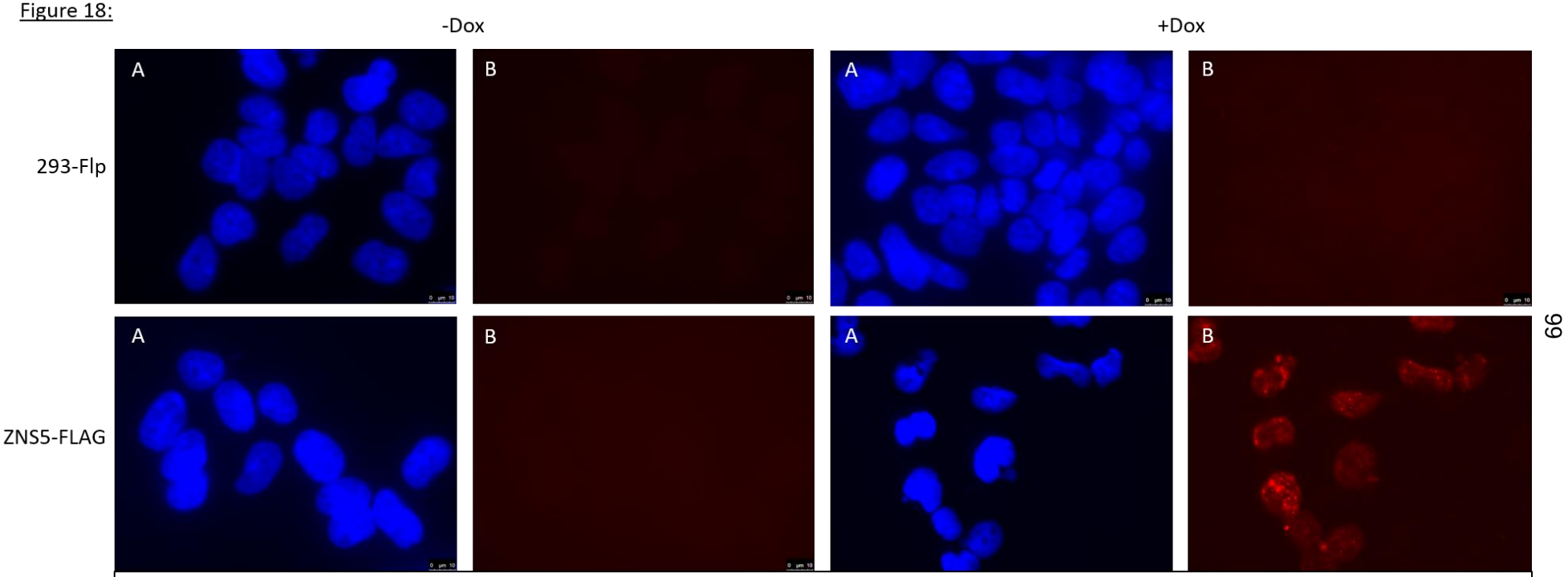

Figure 18: *IFA showing the regulation of ZNS5-FLAG expression with doxycycline* – Both 293-Flp and ZNS5-FLAG cells were incubated with/without doxycycline (1µg/ml) for 4 days. After 4 days they were fixed in 4% paraformaldehyde and for IFA with VectaShield® loading media containing DAPI blue. *A* shows nuclear DAPI blue staining of cells, *B* shows fluorescence from anti-mouse-ab (568) after the samples were initially exposed to anti-FLAG antibodies. The images were taken using widefield microscopy at 100x magnification.

The ZNS5-FLAG cell line was therefore determined to be adequately under the control of doxycycline induction and of adequate expression levels to progress with.

3.4: Production and testing of a ZNS5-BirA cell line

The same 293-Flp-in system was used to produce the ZNS5-BirA cell line. As described in 3.1, the ZNS5-BirA plasmid construct was created using the ZNS5-FLAG plasmid and a BirA-tag containing plasmid such that the ZNS5 gene was inserted upstream of the BirA-tag. The resultant plasmid DNA was propagated as a plasmid midi-prep and examined by restriction enzyme digest. The resulting DNA digest revealed two bands that have the predicted sizes of the ZNS5 gene insert and the BirA plasmid backbone (Figure 19).

Figure 19:

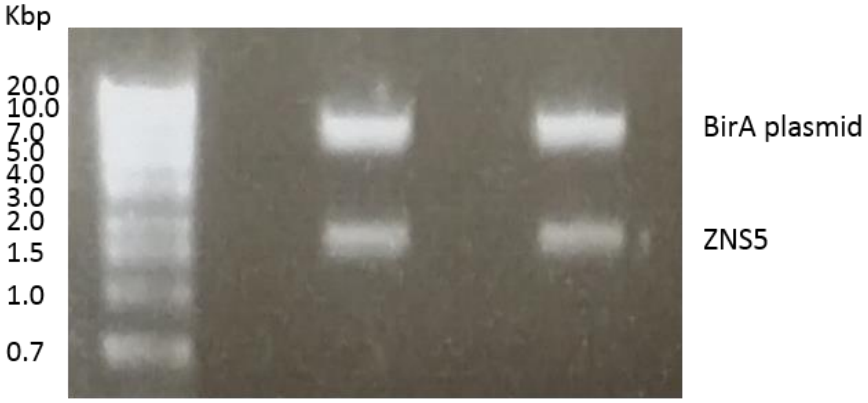

Figure 19: *DNA electrophoresis of post-midi-prep ZNS5-BirA plasmids after double RE digest* – The post-midi-prep ZNS5-BirA plasmids were digested with HindIII and BamHI restriction enzymes for 1hr at 37^{o}C before being run on a 1% agarose gel at 100V for 1hr alongside GeneRuler™ 1kb Plus DNA Ladder (Thermo Scientific). The resulting gel was then imaged using a UVP Biodoc-IT™ System Ultraviolet Transilluminator UV spectrometer.

The sequence of the codon optimized ZNS5 gene and the BirA gene were then used to design primers for sequencing (shown in Table 2). DNA sequence analysis of the plasmid midipreps confirmed that the ZNS5 gene and BirA sequence had acquired no mutations during the cloning or midi-prep procedures. The BLAST nucleotide comparison of the results with the original gene sequence showed that there were no mutations (data not shown).

The ZNS5-BirA plasmid was then used to produce a 293-Flp cell line stably expressing ZNS5-BirA as described in Sections 2.2.8 and 3.3 and shown in Figure 20. Initially 293-Flp cells were transfected in parallel for transient and stable gene expression analysis. 293-Flp cells were transfected according to the protocol described in 2.2.7 and then were seeded on coverslips and either exposed to doxycycline for 4 days or not exposed at all before being fixed and treated with the corresponding antibodies. Cells transiently expressing ZNS5-BirA were used as a control for the purpose of testing the anti-BirA (Aviva Systems Biology OARA01977) (Figure 21) and anti-ZNS5 antibodies (GeneTex GTX133312) (Figure 22) and confirming production of the ZNS5-BirA protein. 293-Flp cells exposed to the same conditions were used as a control for non-specific

binding of the antibodies. For stable cell line production, the initially transfected cells were detached and reseeded into petri dishes after 2 days followed by incubations with hygromycin to select for stably transformed cells. Three colonies were successfully isolated and cultured to test for stable expression of ZNS5-BirA; called clone 3 (C3), clone 4 (C4), and clone 5 (C5).

Figure 20:

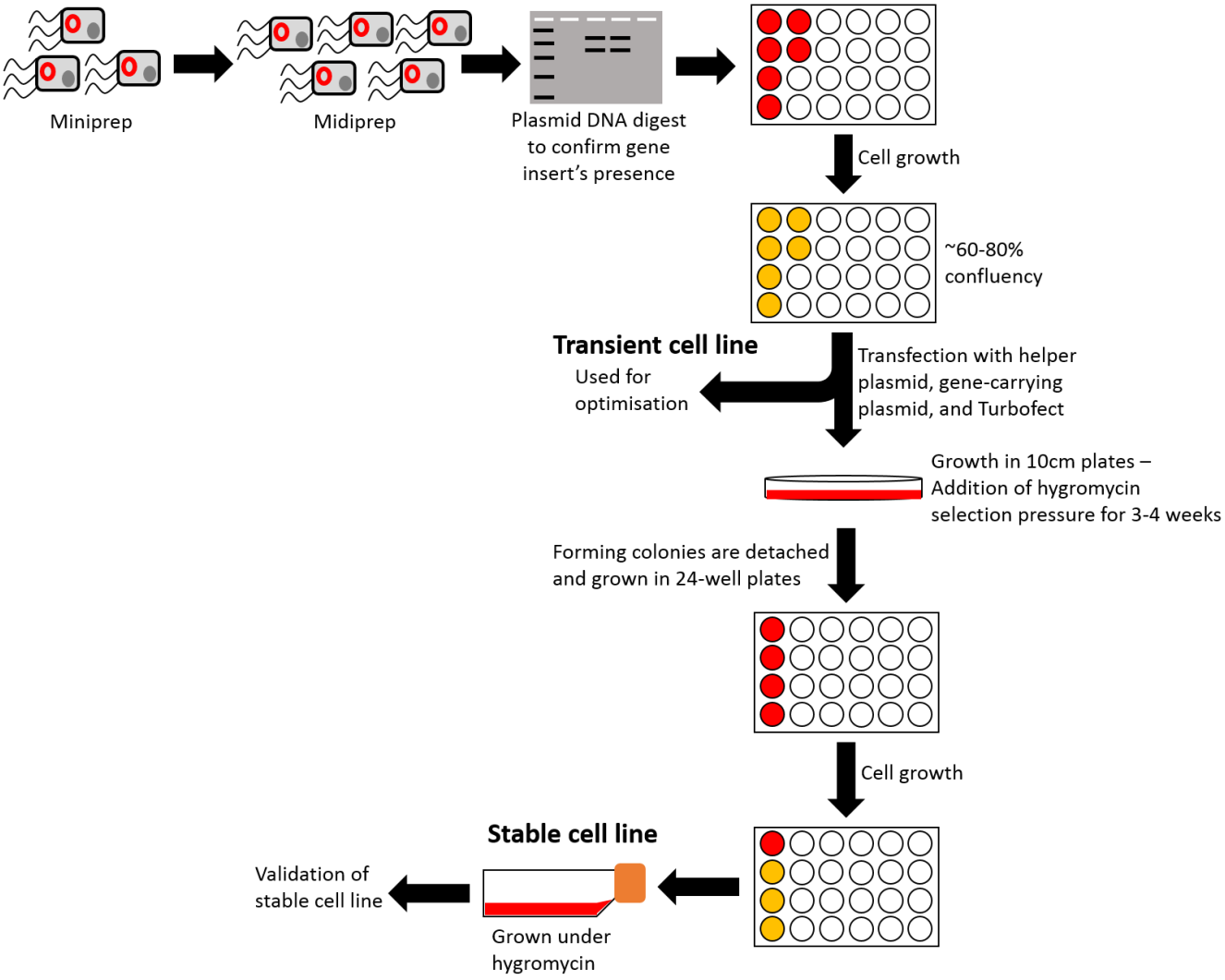

Figure 20: *Generation of transient/stable cell lines* – Outline of the different stages taken in producing the ZNS5-FLAG and ZNS5-BirA transient and stable cell lines and the extra steps taken to produce a stable, transfected population

Figure 21:

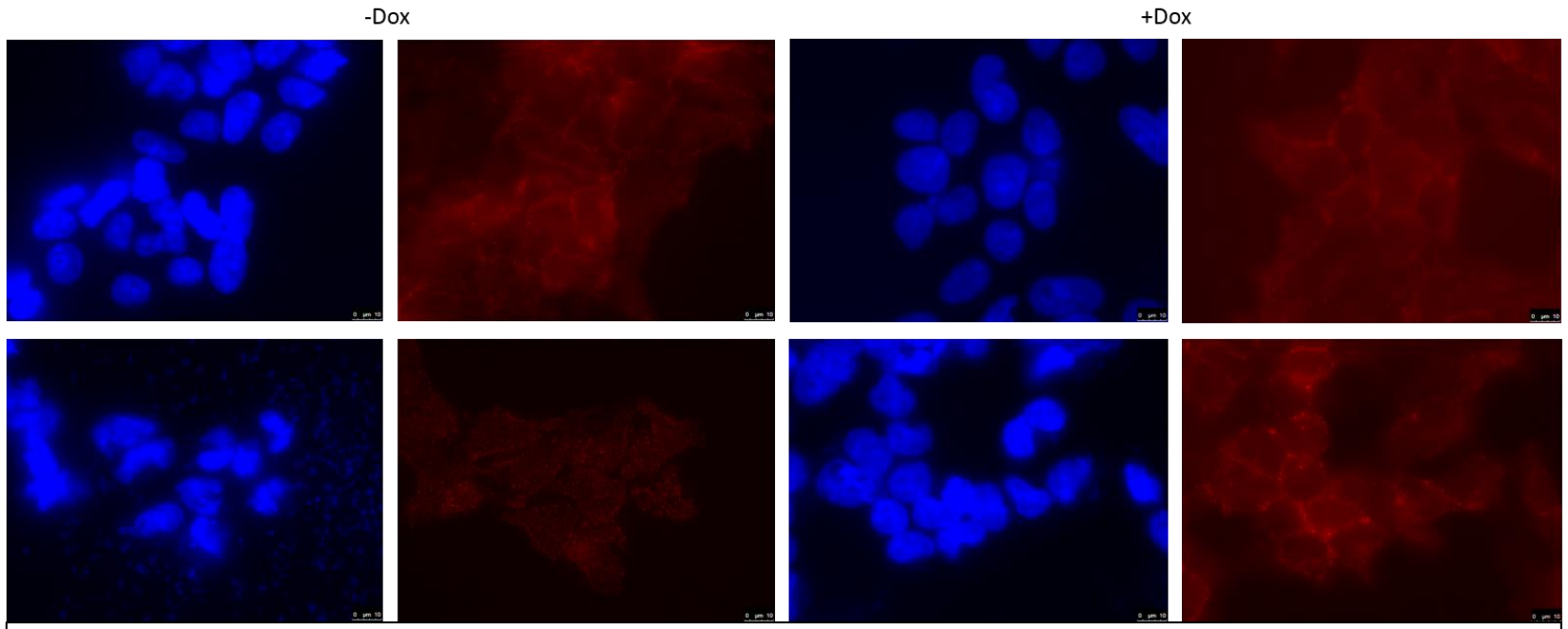

Figure 21: *IFA showing the specificity of the BirA antibody on ZNS5-BirA-expressing cells* - HEK293-Flp-In™ were transiently transfected with the ZNS5-BirA plasmid and then incubated with/without doxycycline (1µg/ml) for 4 days and compared to HEK293-Flp-In™ incubated with/without doxycycline (1µg/ml) for 4 days. All samples were used in IFA with VectaShield® containing DAPI blue nuclear stain. The subsequent images were taken with a widefield microscope at 100x magnification examining; *A*) nuclear DAPI blue staining, and *B*) anti-rabbit-ab (568) antibody/anti-BirA antibody.

Figure 22:

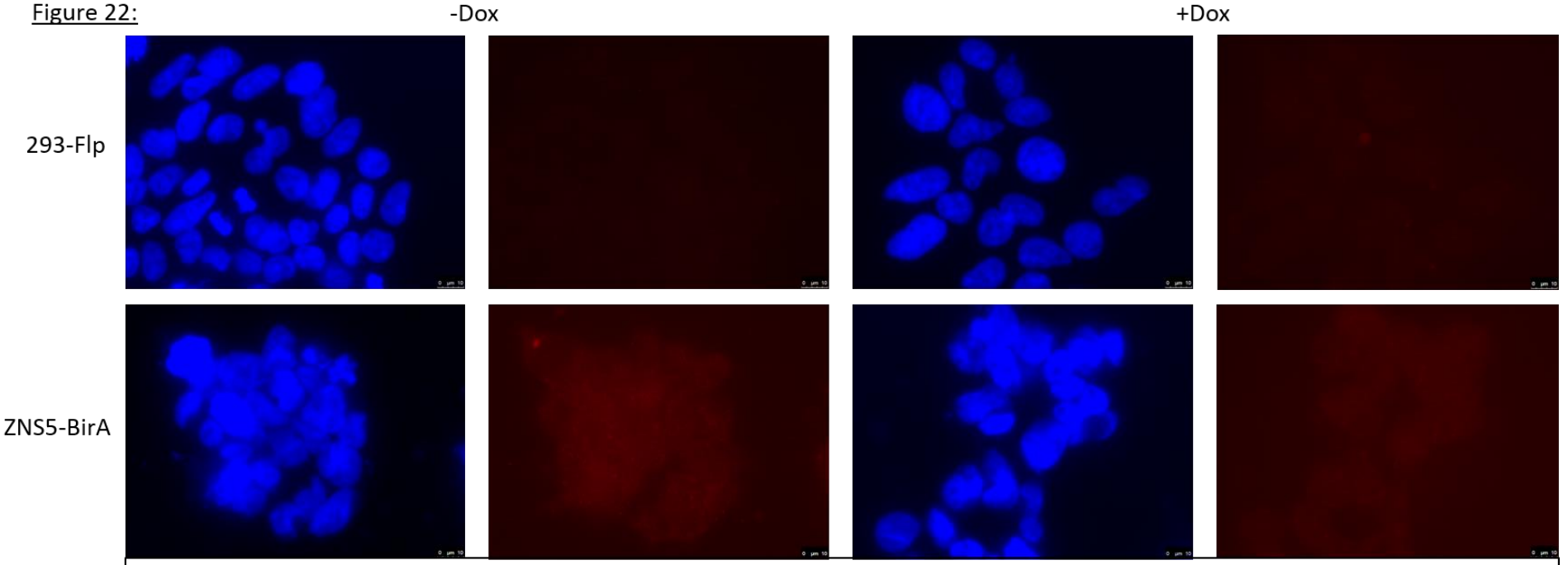

Figure 22: *IFA showing the specificity of the ZNS5 antibody on ZNS5-BirA-expressing cells* – HEK293-Flp-In™ were transiently transfected with the ZNS5-BirA plasmid and then incubated with/without doxycycline (1µg/ml) for 4 days and compared to HEK293-Flp-In™ incubated with/without doxycycline (1µg/ml) for 4 days. All samples were used in IFA with VectaShield® containing DAPI blue nuclear stain. The subsequent images were taken with a widefield microscope at 100x magnification examining; *A*) nuclear DAPI blue staining and, *B*) anti-rabbit-ab (568) antibody/anti-ZNS5 antibody

As shown in Figure 21, the anti-BirA antibody did not specifically detect ZNS5-BirA using IFA analysis, as similar levels of fluorescence were observed in 293-Flp cells which did not have the ZNS5-BirA construct present. An anti-ZNS5 antibody was then used for IFA analysis (Figure 22) which showed more potential, only binding to the ZNS5 expressing cells, but the fluorescence was very weak, so determining whether the antibody specifically recognised ZNS5-BirA was difficult to judge. There also appeared to be some non-specific binding of the antibody to cells even without doxycycline induction.

To investigate further whether the anti-ZNS5 antibody specifically recognised the ZNS5 protein and whether the protein was produced, a western blot was performed using cell lysates prepared from C3/C4/C5 and cell lysates from the 'day 4' and 'day 6' timepoints from the ZNS5-FLAG doxycycline exposure validation experiment, which had already confirmed the presence of the ZNS5 protein. As shown in Figure 23, the anti-ZNS5 antibody failed to detect a band either at the size expected for the ZNS5-BirA protein or the ZNS5-FLAG protein, previously easily detected using the anti-FLAG antibody.

Figure 23:

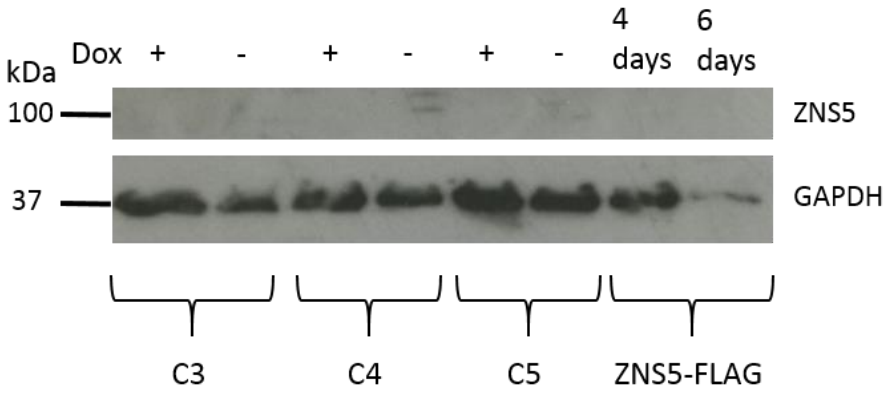

Figure 23: *Western blot showing the efficacy of the ZNS5 antibody* - Cell lysates from the C3/C4/C5 with/without 4 days doxycycline (1µg/ml) incubation and cell lysates from the ZNS5-FLAG cell line from the previous experiment (Figure 17) were used in a western blot performed with anti-ZNS5 and anti-GAPDH antibody. The resulting blot was developed in a Konica film developer.

As a result of having no capable antibodies, a functional assay was performed in an attempt to detect ZNS5-BirA protein expression. All three ZNS5-BirA cell clones were cultured with and without doxycycline for 4d alongside 293-Flp cells. Biotin was added on the third day and 18h after biotin addition, the cells were lysed. The protein concentrations in the cell lysates were quantified by BCA assay and then equal quantities of protein were used for a fluorescent western blot. As biotin was the target, green-fluorescent-conjugated Streptavidin was used and as skim milk-powder can contain biotin, the samples were blocked using 5% BSA (in PBS containing 0.02% SDS and 0.1% Tween). After 4d of doxycycline (1µg/ml) exposure and 18h of biotin (2µg/ml)

exposure, all three ZNS5-BirA cell clones have high levels of biotinylated proteins which was not observed in any samples that lacked biotin, doxycycline, or both treatments (Figure 24).

Figure 24:

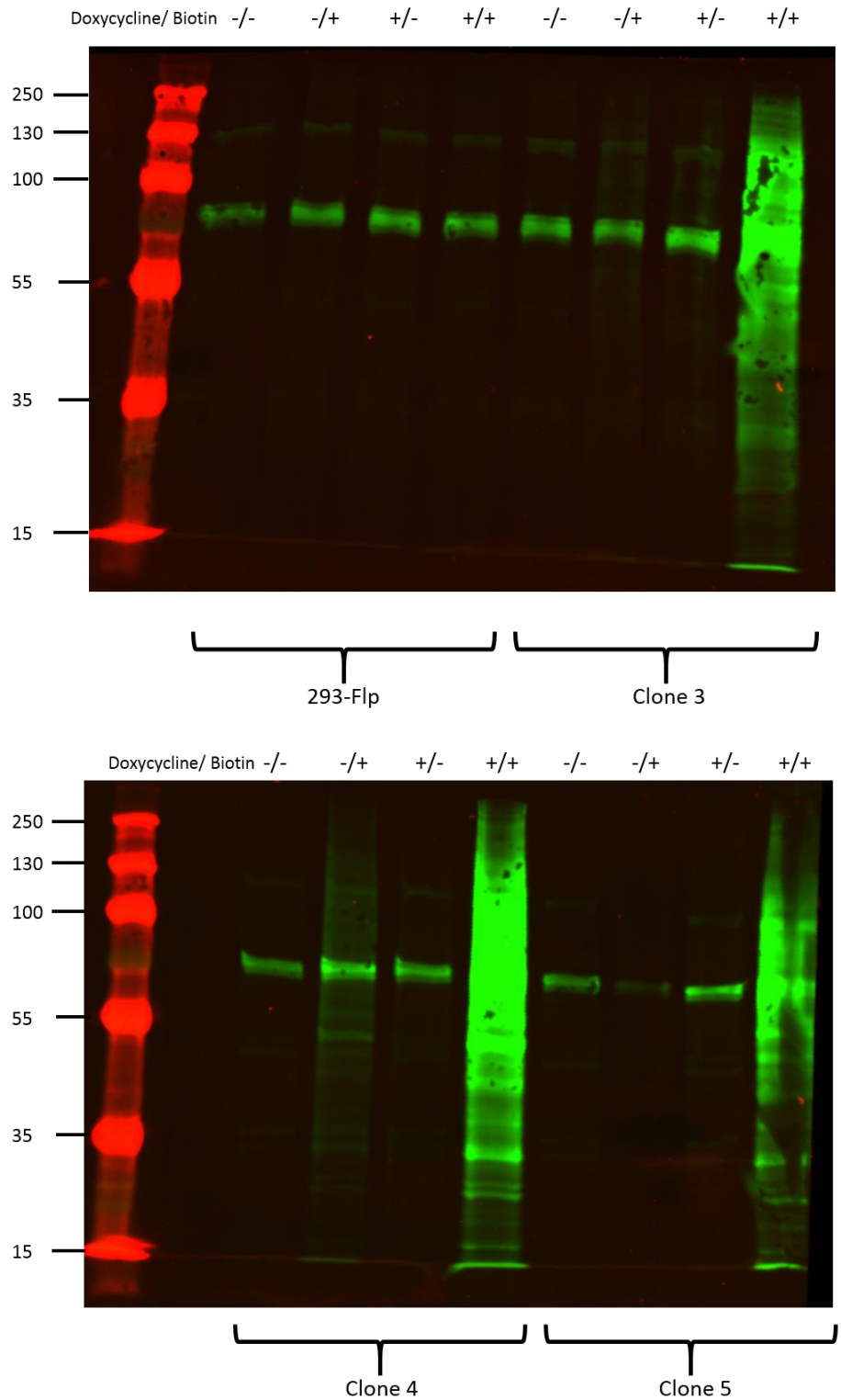

Figure 24 : *Fluorescent WB on ZNS5-BirA cells probing for biotin* – C3, C4, and C5 ZNS5-BirA cell lines were incubated for 4 days with doxycycline (1µg/ml) and for 18 hours with biotin (2µg/ml) (+/+), incubated for 4 days with doxycycline and no biotin (+/-), incubated with no doxycycline but for 18 hours with biotin (-/+), or with neither doxycycline or biotin (-/-). The cells were lysed and a BCA assay was performed to ensure equal loading and then the subsequent western blot was performed with Streptavidin Alexa Fluor™ 488 conjugate. The resulting blot was then imaged in a Li-Cor Odyssey imaging machine.

In addition, only a single protein was detected in lysates from the parental cell line which was also present in lysates derived from the ZNS5-BirA cells. RT-PCR was also performed on RNA extracted from C3, C4, and C5 cells that had been cultured with/without doxycycline (1µg/ml) for 4d using primers specific for ZNS5-BirA RNA (shown in table 2 in 2.2.4) to confirm the transcription of the ZNS5-BirA gene. The process was performed using primers that would amplify two overlapping ~1.9kb fragments. The results (Figure 25) show the amplification, and therefore presence, of ZNS5-BirA RNA in all samples induced with doxycycline. There was also some level of ZNS5-BirA RNA in samples not induced with doxycycline and therefore amplification of the created cDNA, suggesting some low-level transcription occurred even without doxycycline.

Figure 25:

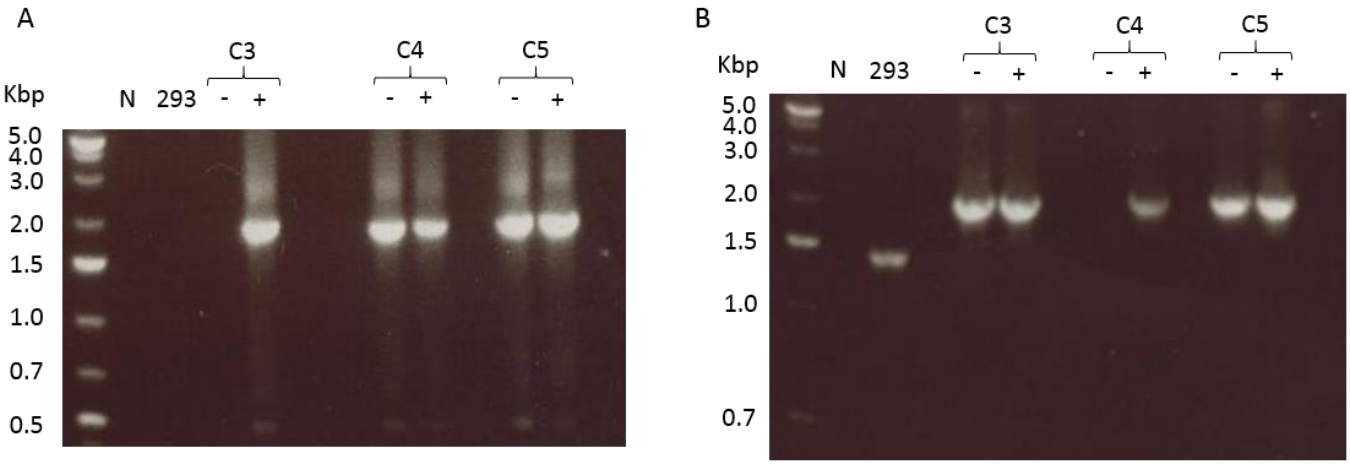

Figure 25: *RT-PCR products of C3/C4/C5 with/without 4 days doxycycline induction using ZNS5-BirA primers* – C3/C4/C5 cells were cultured for 4 days with(+)/without(-) doxycycline before RNA extraction and subsequent RT-PCR using primers designed for the ZNS5-BirA gene (primers described in 2.2.4). The RT-PCR was carried out in two ~1.9kb reactions; *A*) prStart + prMidRev, and *B*) prMid + prEnd. A 'no polymerase'(N) reaction was included as well as a negative control of HEK-293(293) under the induction of 4 days doxycycline.

The products were sent to Eurofins sequencing and the results confirmed the expression of ZNS5-BirA without mutations (Appendix B). The results suggested that all three cell lines stably expressed ZNS5-BirA in a relatively inducible fashion.

3.5: Pulldowns

3.5.1: NS5-FLAG pulldown

Using the previously produced stable ZNS5-FLAG cell line, 293-Flp cells, and previously constructed D2NS5-FLAG and D4NS5-FLAG-expressing cells, 3x T225 flasks of cells were grown and incubated with doxycycline (1µg/ml) for 4d and then lysed in mild, non-denaturing lysis buffer to create large pool of proteins without interfering with protein-protein interactions. Using anti-FLAG antibody-coated magnetic beads, it was possible to extract the FLAG tagged NS5 proteins by affinity purification as well as any proteins that bound to them (protocol described in further detail in

2.3.8). In this way, the cellular interactome of the respective NS5 proteins could be isolated for further analysis. A sample taken from each lysate at the start of the procedure was compared with a small sample of the proteins bound to the anti-FLAG magnetic beads at the end of the procedure using a western blot to determine if there had been an increase in concentration of the NS5-FLAG proteins as would be expected of this procedure. Figure 26 shows an increase in concentration of protein detected by anti-FLAG antibody at ~105kDa, the expected size of the NS5-FLAG proteins. A band of a similar size was not observed in the 293-Flp control cells and so it was decided that the pulldown was a success.

Figure 26:

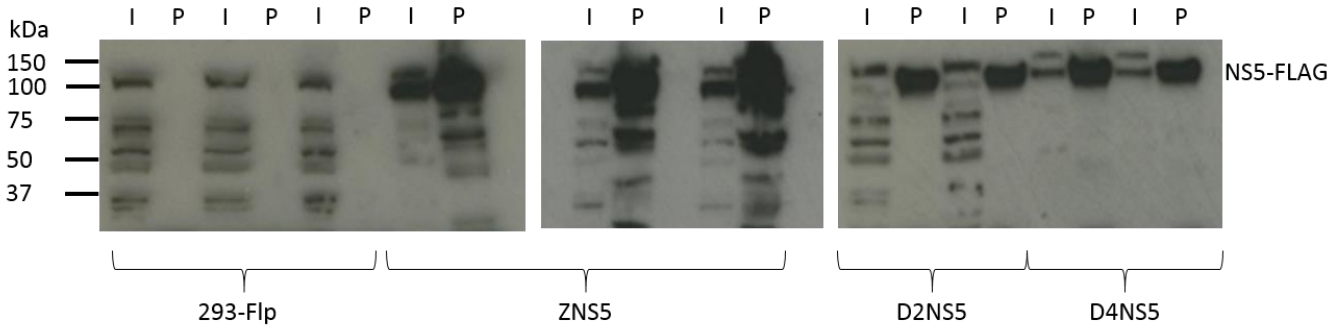

Figure 26: *Western blot showing pre- and post- FLAG pulldown samples* - Three replicates of both HEK293 Flp-In™ and ZNS5-FLAG cell lines, and two replicates of D2NS5 and D4NS5 cell lines were incubated with doxycycline (1µg/ml) for 4 days while growing to high confluency in T225 flasks. The cells were lysed and a small sample was taken from the initial lysate with the rest undergoing FLAG pulldown. A small sample was taken after the pulldown and the pre-(*I*) and post-(*P*) pulldown samples were then compared by western blot using anti-FLAG antibody. The film was then developed using a Konica film developer

The remainder of the FLAG magnetic beads that retained the captured NS5-FLAG proteins and their cellular interaction partners were then sent to the University of Bristol Proteomics facility for tryptic digestion of the proteins bound to the beads, differential 10-plex TMT labelling and LC-MS/MS analysis.

3.5.2: Streptavidin pulldown

Whole cell lysate pulldowns using Streptavidin coated beads had not been done previously in the laboratory, therefore a protocol was developed based on the whole cell lysate/ FLAG tag affinity purification protocol and a previously reported protocol for the isolation of biotin-labelled proteins using Streptavidin coated beads (204). C3, C4, and C5 and three replicates of 293 Flp-In™ cells were grown in 3x T225 flasks and incubated with doxycycline (1µg/ml) for 4d and biotin (2µg/ml) for 18h to promote proximity labelling from the BirA tag. The cells were lysed in the same mild, non-denaturing lysis buffer as the FLAG pulldown and the lysate was then processed according to the pulldown procedure outlined in 2.3.10 using the Streptavidin-coated beads. Proteins not labelled with biotin were removed using strong detergent and high salt conditions to give a final sample that contains only labelled proteins. A sample taken from the initial lysate of

each sample was compared to the same sample post-pulldown on a fluorescent western blot (protocol in 2.3.6) using fluorescent Streptavidin as a biotin probe. The results, shown in Figure 27, showed that C4 in particular pulled down numerous biotinylated proteins and best reflected the biotinylation show in figure 24. C3 had enriched some biotinylated proteins though was clearly less successful than C4 and C5 showed no clear signs of having been successful.

Figure 27:

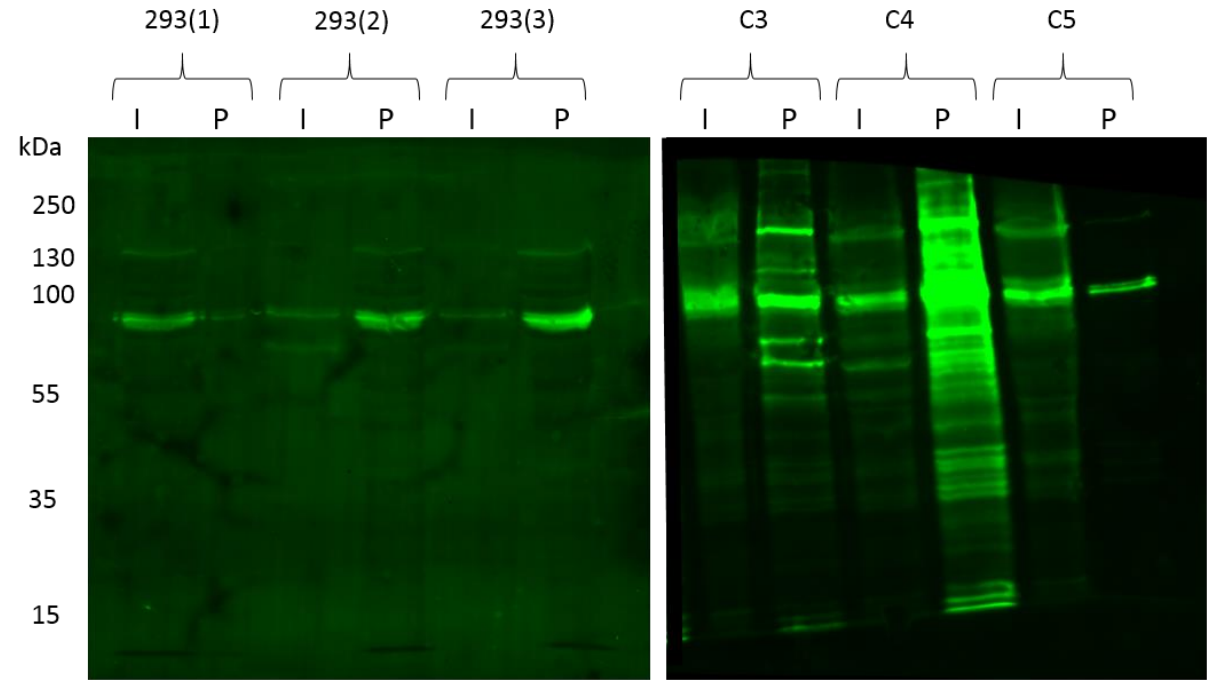

Figure 27: *Fluorescent western blot showing pre- and post- Streptavidin pulldown samples* – 3x T225 flasks of HEK293 Flp-In™ cells 1x T225 flask of each of C3/C4/C5 and were incubated for 4 days with doxycycline (1µg/ml). All were incubated with biotin (2µg/ml) for 18 hours. They were then lysed and a small sample (*I*) was kept for analysis before the streptavidin pulldown was performed on the remaining lysate. A further sample after the pulldown (*P*) was then compared to the pre-pulldown sample on a fluorescent western blot with Streptavidin Alexa Fluor™ 488.

It was decided to send all of the samples to be analysed by TMT-MS/MS and remove C5 from the subsequent bioinformatic analysis if the results showed the pulldown to have failed to enrich biotinylated proteins compared with C3 and C4.

3.6: FLAG TMT-MS/MS and bioinformatic programme comparison

3.6.1: TMT-MS/MS

TMT labelling allows for up to 11 samples to be analysed in a single LC-MS/MS run. Each peptide population derived from a protein lysate by proteolytic digestion is labelled with a specific mass tag that uses different isotopes of oxygen, carbon and nitrogen in different positions within the tag molecule so that while the overall mass of the tag remains the same, a cleavable portion of

the tag (known as the mass reporter) allows identification of each tag when analysed by LC-MS/MS.

The process of sample set-up and TMT-LC-MS/MS is shown in Figure 28. The proteins bound to the FLAG beads are eluted from the beads and then digested. The peptide fragments are then labelled with a TMT tag, each sample receiving a different tag. All the peptides from every sample are then pooled and re-separated using column liquid chromatography before being analysed by MS/MS to detect the fragment identities. As the tags only add ~130 Da onto the peptide fragments, the chromatography is not greatly affected by their addition. The TMT tags also contain a 'cleavage linker' that breaks during a specific stage of the MS/MS analysis of the peptide fragment, allowing for the release of the mass reporter. The mixed mass reporters are detected at the same time as the peptide fragments they were linked to and due to the altered locations of the isotopes within each tag's structure, each tag has a distinct mass which can be detected by MS/MS. The quantity of each tag can therefore be quantified and compared to each other sample for each individual peptide fragment. The raw data that is returned can be analysed using a number of different proteomic analysis programmes which use protein search databases based on the species of the host cell used or a custom list, to match the peptide fragment data to prospective proteins, giving a final readout of the proteins that are predicted to have been detected and the relative quantities at which they are present in each sample.

Figure 28:

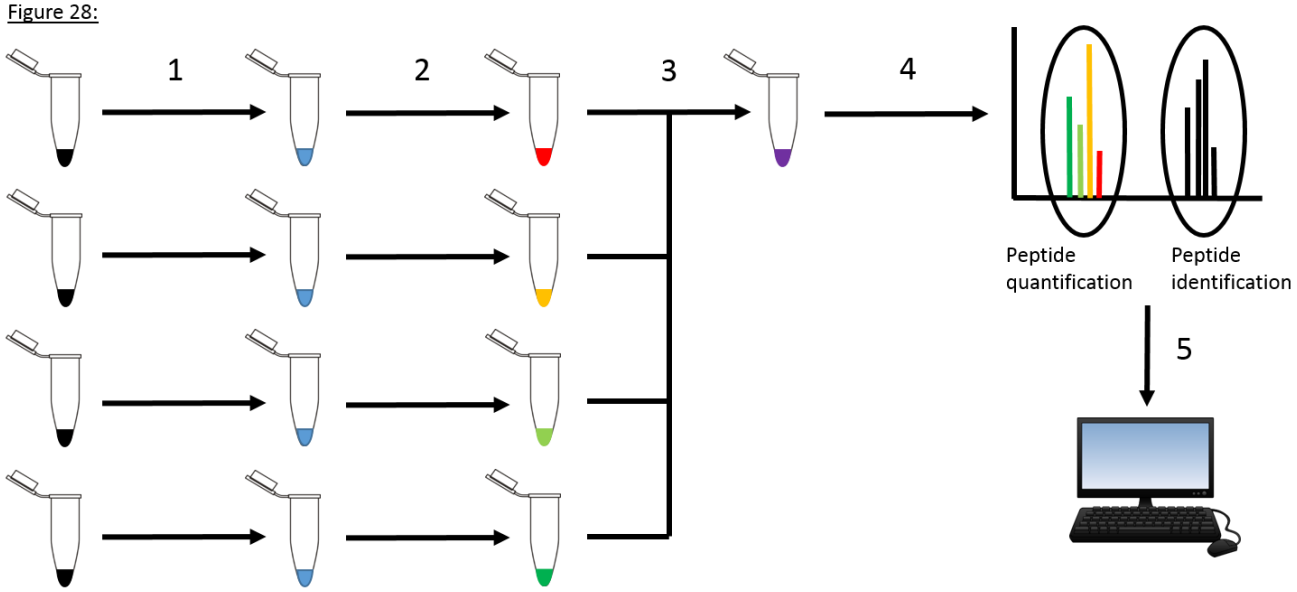

Figure 28: *Sample preparation and data generation for TMT-MS/MS proteomics* – 1) The proteins are removed from the beads and the beads are discarded. 2) The proteins are degraded using proteases and then the TMT labelling tags are added which label each peptide in that sample with a unique tag for that sample. Up to 10 different samples can be used in one run of the MS/MS as there are 10 unique identifiable tags currently commercially available. 3) The samples are pooled together and then 4) run through a liquid chromatography column and mass spectrometer. The results give both the MS results for identification of each peptide as well as peaks for the different tags that were attached to the peptide fragment. As the samples were all pooled together, all the tags are detected at the same time and the comparison of the tag peaks for each individual peptide gives the relative quantity of the peptide in each sample. 5) The results are then passed through proteomics programmes like MaxQuant or Proteome Discoverer which then predicts which protein sourced the peptide fragments to build a comprehensive database of which proteins are present within the samples and how their quantities varied between samples.

3.6.2: FLAG pulldown bioinformatics and proteomics programme comparison

The NS5-FLAG pulldown data was analysed using the proteomics bioinformatics software MaxQuant (207) using two false discovery rates (FDR) of 0.01 and 0.05, as well as PDv1.4 (at its default settings) to identify and quantify the proteins that bound to the FLAG beads. The datasets were then filtered to identify the protein hits that were most reliable and common to all three data sets. The processed data was then analysed statistically using an SPSS based programme, Perseus v1.5.0.15 (208), developed specifically for proteomic data analysis. Using Perseus, protein hits that were considered to be contaminants, reverse/mirror proteins, or those that were identified only by a single site were removed from the datasets. A log2 transformation was performed on the raw values which were then normalised by subtraction of the median value from the dataset for each condition, to ensure a normal distribution around the median value as shown by the histograms in Figures 29-31.

Figure 29:

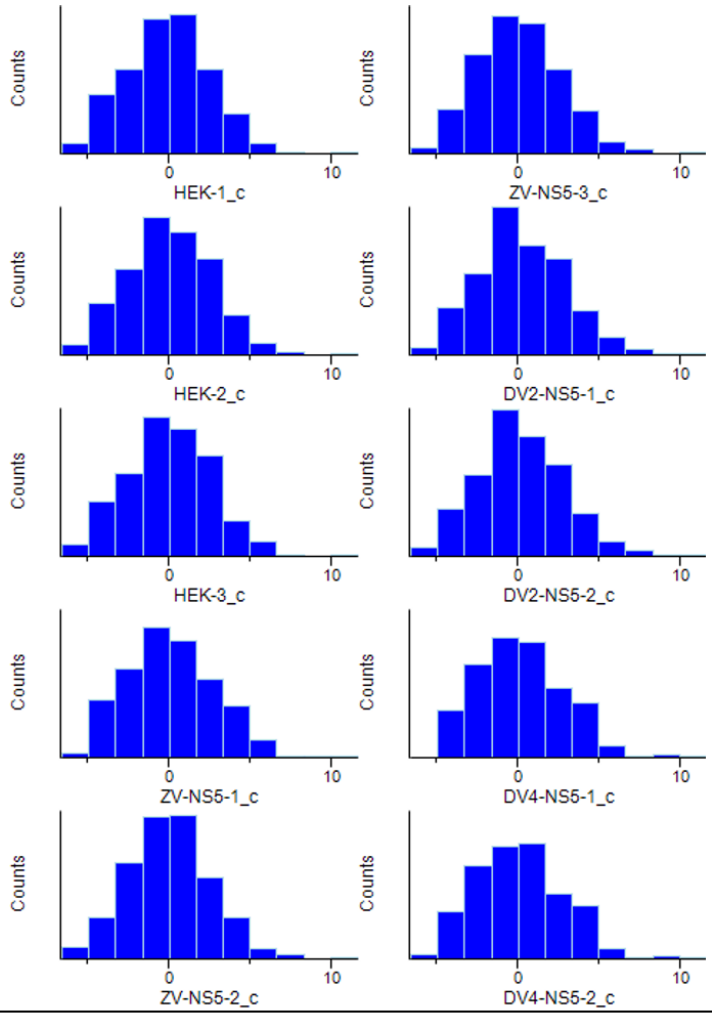

Figure 29 : *Histograms showing the distribution of protein concentrations of each sample from the 0.01 FDR MaxQuant-processed TMT-MS/MS data* – Proteomic data from the MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through MaxQuant at a FDR of 0.01 before being processed in Perseus. A log2 transformation was performed on the protein counts and the data was then normalised using median subtraction to give the shown histograms

Figure 30:

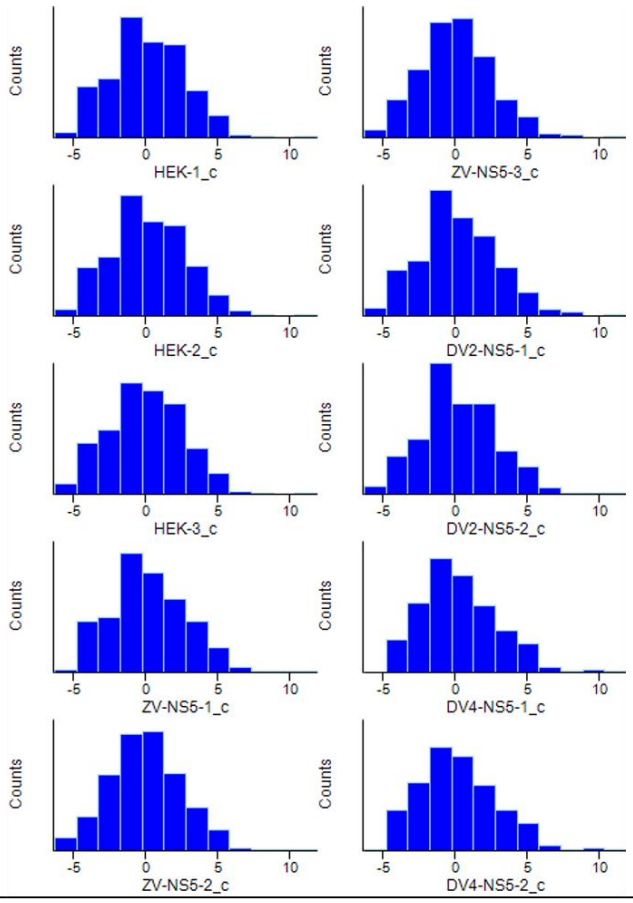

Figure 30: *Histograms showing the distribution of protein concentrations of each sample from the 0.05 FDR MaxQuant-processed TMT-MS/MS data* - Proteomic data from MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through MaxQuant at a FDR of 0.05 before being processed in Perseus. A log2 transformation was performed on the protein counts and the data was then normalised using median subtraction to give the shown histograms

Figure 31:

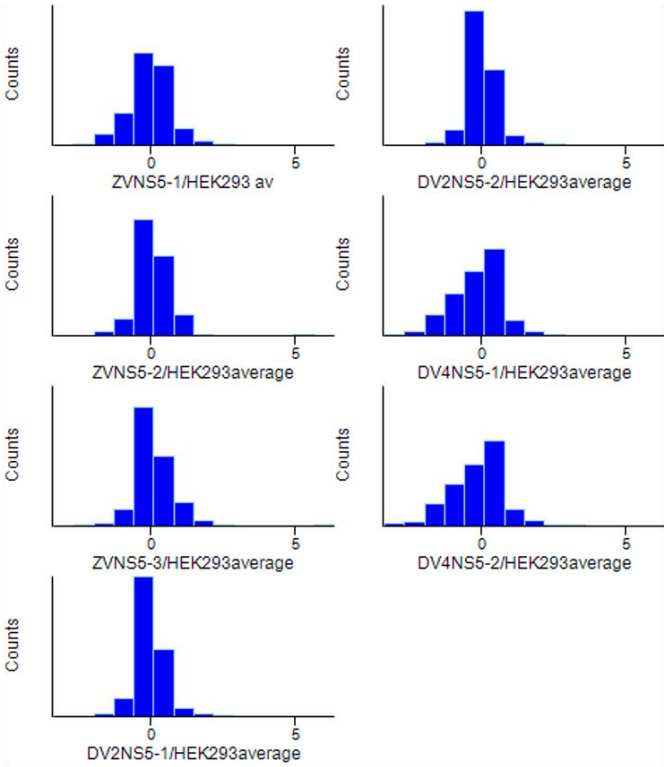

Figure 31: *Histograms showing the distribution of protein concentrations of each sample from the Proteome Discoverer-processed TMT-MS/MS data* - Proteomic data from MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through Proteome Discoverer at its default FDR before being processed in Perseus. A log2 transformation was performed on the protein ratios and the data was then normalised using median subtraction to give the shown histograms

Using Perseus, Pearson's Correlation analysis was performed on the replicate samples within the same category (293-Flp/ZNS5/D2NS5/D4NS5) to determine how closely the replicates of each set correlated with each other with a scatterplot generated as a visual representation of this similarity. As shown in Figures 32-34, the replicates were highly comparable for every dataset, suggesting the results in each were reproducible and reliable however PD showed a lower level of correlation between the same datasets than was determined by MaxQuant at both FDRs.

Figure 32:

Figure 32: *Scatterplot graphs showing the correlation between replicates of FLAG pulldown samples from the 0.01FDR MaxQuant-processed TMT-MS/MS data* - Proteomic data from the MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through MaxQuant at a FDR of 0.01 before being processed in Perseus where scatterplots were generated for replicates of each cell line along with the Pearson's correlation value in the top left of each individual scatterplot

Figure 33:

Figure 33: *Scatterplot graphs showing the correlation between replicates of FLAG pulldown samples from the 0.05FDR MaxQuant-processed TMT-MS/MS data* - Proteomic data from the MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through MaxQuant at a FDR of 0.05 before being processed in Perseus where scatterplots were generated for replicates of each cell line along with the Pearson's correlation value in the top left of each individual scatterplot

Figure 34:

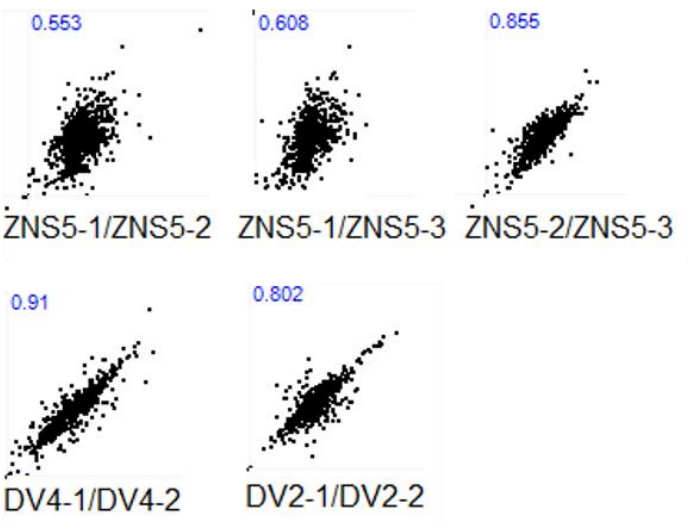

Figure 34: *Scatterplot graphs showing the correlation between replicates of FLAG pulldown samples from the Proteome Discoverer-processed TMT-MS/MS data* - Proteomic data from the MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through Proteome Discoverer at its default FDR before being processed in Perseus where scatterplots were generated for replicates of each cell line along with the Pearson's correlation value in the top left of each individual scatterplot

The output from the MaxQuant analysis provided raw quantitation values for each protein identified and so the data from both the 0.01 and 0.05 FDR analyses could be grouped into categories for two sample t-tests so that each category (ZNS5/D2NS5/D4NS5) could be compared against the control category of 293-Flp. Perseus automatically averaged the samples within the group and performed a two sample t-test so that a single t-test difference value was given as well as a log2 value for that t-test.

PDv1.4 analysis produced ratios comparing each sample to a single chosen control sample, therefore the data of all the expression values of the 293-Flp samples had to be averaged first so that a single ratio was given per replicate of ZNS5/D2NS5/D4NS5. The data used for the Perseus analysis were the ZNS5, D2NS5, and D4NS5 replicates, each with a ratio (to the averaged control values) rather than a raw detection value. Relative quantitation values for the 293-Flp control samples were not produced using PD for analysis as it would always have a ratio of 1 compared to itself for every protein identified. Therefore one sample t-tests were used to analyse the ratios comparing against a log2 values of '0' (ie no change compared to control), whereas the MaxQuant analysis produced abundance values for all conditions which could be used for 2 sided t-test analysis (ie condition vs control).

The resulting datasets were then exported into Microsoft Office Excel to be analysed. The processed data is supplied on the CD in Appendix C. Table 5 summarises the differences between the PD and MaxQuant outputs as well as the difference using different FDRs for both MaxQuant datasets.

A log2 t-test difference of >0.58 was chosen as the lower boundary of our classification of increased protein binding as it equates to roughly a 1.5 fold increase between our sample pulldowns and the 293-Flp control pulldowns. However, proteins with a log2 t-test difference between 0.5-0.58 were also examined in the final data in case of mitigating factors that may have affected t-test values. The p values were given as –log10 p values and so >1.3 represented a p value of <0.05. The ideal protein targets therefore had a log2 t-test difference >0.58 and a –log10 p value of >1.3.

Table 5: *Comparison of raw datasets from the FLAG pulldown analysis using Proteome Discoverer and MaxQuant with 0.01 and 0.05 FDRs.*

Dataset	No. of proteins detected	No, of proteins with >1 peptide	No. of proteins log2 t-test >0.58 for ZNS5	No. of proteins –log10 p value >1.3 for ZNS5	No. of proteins log2 t-test >0.58 for D2NS5	No. of proteins –log10 p value >1.3 for D2NS5	No. of proteins log2 t-test >0.58 for D4NS5	No. of proteins –log10 p value >1.3 for D4NS5
MaxQuant - 0.01 FDR	544	316	117	140	104	176	135	292
MaxQuant - 0.05 FDR	669	329	113	162	111	203	173	343
PD	856	856	114	135	78	155	180	371

High confidence protein hits within each dataset were separated for each NS5 interactors list by isolating protein hits that had a >0.58 log2 t-test difference. Proteins down to 0.5 were also considered if there were reasons to include it such as a single poor replicate or a failed occurrence in a replicate. Proteins were then determined to be reliable or not based on a –log10 p value >1.3. Due to the nature of variability within proteomic data and in the interest of generating the largest pool of potential interactors, proteins with a –log10 p value <1.3 were manually reviewed regarding their counts, unique peptides, coverage scores, and their individual replicate values to determine the cause of the lower p value. In numerous cases, it was found to be the result of the identification of a protein using a single replicate in either the control or pulldown sample and in these cases, and where a protein was confidently identified by high coverage, unique peptide number, and count value, it was also included into the dataset. Inclusion of these proteins did not affect the enrichment of particular pathways shown in STRING and DAVID analysis (data not shown). This final protein list was considered to be the ‘high confidence proteins’ for the general interactomic analysis. Table 6 shows a summary of the number of high confidence proteins for each sample.

Table 6: *Comparison of the quantity of high confidence proteins of NS5-FLAG from Proteome Discoverer and MaxQuant at 0.01 and 0.05FDR*

Dataset	No. of high confidence proteins for ZNS5	No. of high confidence proteins for D2NS5	No. of high confidence proteins for D4NS5
MaxQuant - 0.01 FDR	123	116	138
MaxQuant - 0.05 FDR	112	105	175
PD	123	95	153

The analysis therefore produced lists of high confidence protein hits that could then be compared between the datasets and between the different flaviviral NS5 proteins. A summary of the common high confidence proteins between the different samples in each dataset is shown below in table 7.

Table 7: *Comparison of the quantity of common high confidence protein interactors of NS5-FLAG identified by the Proteome Discoverer and MaxQuant at 0.01 and 0.05 FDR analyses.*

Datasets	No. common high confidence proteins for ZNS5	No. common high confidence proteins for D2NS5	No. common high confidence proteins for D4NS5
0.01 and 0.05	83	85	120
0.01 and PD	46	46	60
0.05 and PD	40	47	70
0.01, 0.05, and PD	39	40	55

It was also possible to compare the same samples from different datasets to see which hits were consistently being discovered using different programmes and different FDRs. These proteins were considered to be the most reliable hits. There were 76 proteins identified to be present in all three datasets that had a >0.58 log2 fold change in at least one of the NS5 sample sets (ZNS5/D2NS5/D4NS5) compared to the control. It was these proteins that were most suited for validation as they were present in all datasets and identified as binding at increased levels to at least one of the NS5 proteins. These proteins are shown in Table 8 with the log2 t-test difference provided according to 0.01FDR MaxQuant.

Table 8: *Common high confidence proteins from the NS5-FLAG pulldowns*

Protein name	Accession number	log2 ZNS5 t-test difference	log2 D2NS5 t-test difference	log2 D4NS5 t-test difference
Cysteine-rich PDZ-binding protein	Q9P021	2.51	0.13	2.44
40S ribosomal protein S10;Putative 40S ribosomal protein S10-like	P46783	1.91	2.38	2.63
Reticulocalbin-2	H0YL43	1.79	0.78	1.84
28 kDa heat- and acid-stable phosphoprotein	Q13442	1.66	2.78	1.45
40S ribosomal protein S20	P60866	1.65	2.49	2.55
40S ribosomal protein S25	P62851	1.56	2.47	2.40
40S ribosomal protein S29	P62273	1.55	2.53	2.26
Slit homolog 2 protein	A0A087WYV5	2.44	2.09	1.94
Cleft lip and palate transmembrane protein 1	K7ERL5	1.45	0.67	0.67
ATP synthase subunit gamma (fragment)	Q8TAS0	1.77	1.20	1.47
Peptidyl-prolyl cis-trans isomerase FKBP3	Q00688	1.37	1.56	0.98
Cyclin-dependent kinase 1	P06493	1.29	0.52	0.31
ATP synthase subunit alpha, mitochondrial	K7ENP3	1.29	0.84	0.50
E3 ubiquitin-protein ligase listerin	H7BYG8	1.24	-0.02	-0.14
Eukaryotic translation initiation factor 3 subunit B	C9JZG1	1.22	0.90	0.86
Protein FAM133B	Q5BKY9	1.14	1.38	0.26

Eukaryotic translation initiation factor 5B	O60841	1.11	0.47	1.34
40S ribosomal protein S19	P39019	1.19	2.35	2.65
ATP synthase subunit O, mitochondrial	P48047	1.08	0.76	2.00
40S ribosomal protein S18	P62269	1.08	1.18	1.86
40S ribosomal protein S28	P62857	1.18	2.15	1.31
Signal recognition particle receptor subunit beta	H7C4H2	1.03	0.05	0.36
40S ribosomal protein S17-like;40S ribosomal protein S17	H0YN88	0.97	1.23	2.10
ATPase inhibitor, mitochondrial	Q9UII2-3	0.94	0.93	0.85
40S ribosomal protein S30	E9PR30	0.83	1.18	1.59
Protein FRG1	Q14331	0.84	0.94	1.33
Transcription initiation factor TFIID subunit 2	Q6P1X5	0.72	1.84	1.50
Lamin B receptor, isoform CRA_a	A0A024R3R5	0.76	0.65	-1.73
Vimentin	Q5JVS8	0.68	0.40	-0.14
MICOS complex subunit MIC60	Q16891-2	0.68	0.15	-0.70
Histone H2A type 2-C	Q16777	0.79	0.15	0.77
60S ribosomal protein L12	P30050-2	0.65	0.97	2.18
40S ribosomal protein S16	P62249	0.65	0.37	0.43
60S ribosomal protein L9	H0Y9V9	0.63	1.13	2.36
Arginine and glutamate-rich protein 1	Q9NWB6-2	0.62	-0.02	2.039

40S ribosomal protein S7	B5MCP9	0.62	0.61	1.55
PHD finger protein 6	Q5JRC6	0.61	0.57	1.18
High mobility group protein HMGI-C	P52926	0.60	0.51	1.34
Elongation factor Tu, mitochondrial	P49411	0.59	0.95	-0.34
RNA-binding protein 25	P49756	0.57	0.48	0.99
Pre-mRNA 3-end-processing factor FIP1	Q6UN15-3	0.54	0.88	1.36
Pre-mRNA-processing factor 40 homolog A	O75400	0.69	0.59	0.91
Protein BUD31 homolog	C9JNV2	0.51	0.44	0.85
GTP-binding nuclear protein Ran	B5MDF5	0.50	0.66	1.93
Histone H1.2;Histone H1.3	P16403	0.49	0.76	1.18
Emerin	Q5HY57	1.21	0.36	0.98
Plasminogen activator inhibitor 1 RNA-binding protein	Q8NC51-2	0.43	1.08	1.10
Non-histone chromosomal protein HMG-14	A6NL93	0.43	0.03	1.51
40S ribosomal protein S23	P62266	0.57	1.26	-0.02
Protein SREK1IP1	Q8N9Q2	0.40	0.97	-0.04
Histone H1.4	P10412	0.36	0.85	1.00
60S ribosomal protein L37a	P61513	0.47	0.54	1.28
60S ribosomal protein L23a	P62750	0.45	0.34	1.40
Calnexin	D6RAQ8	0.34	0.84	-0.18
60S ribosomal protein L22	K7ERI7	0.32	0.37	1.26

High mobility group protein HMG-I/HMG-Y	P17096	0.14	0.50	1.18
40S ribosomal protein S15	K7EJ78	0.55	0.53	1.18
60S ribosomal protein L32	D3YTB1	0.28	1.55	0.42
40S ribosomal protein S14	P62263	0.25	0.05	1.35
Cytochrome b-c1 complex subunit 2, mitochondrial	H3BRG4	0.31	0.68	1.73
Endothelial differentiation-related factor 1	O60869-2	0.24	1.37	0.55
Histone H4	P62805	0.20	0.13	0.96
Leucine-rich repeat-containing protein 59	Q96AG4	0.18	0.60	0.87
Poly(U)-binding-splicing factor PUF60	Q9UHX1-4	0.13	0.02	0.83
Serine/arginine-rich splicing factor 11	Q05519-2	0.11	-0.11	0.88
Protein DEK	P35659	0.18	0.09	1.19
Cold-inducible RNA-binding protein	Q14011	0.07	0.89	1.46
High mobility group protein B1	P09429	1.16	0.63	0.93
High mobility group protein B2	P26583	0.03	0.57	1.08
High mobility group nucleosome-binding domain-containing protein 4	O00479	-0.01	0.16	1.57
Uncharacterized protein C19orf47	Q8N9M1-3	-0.06	-0.06	1.35
Serine/threonine-protein kinase RIO1	Q9BRS2	-0.07	1.02	-0.22

High mobility group protein B3	E9PES6	-0.39	0.30	1.12
NF-kappa-B-activating protein;NKAP-like protein	Q8N5F7	-0.45	-1.24	0.93
Kinesin-like protein KIF11	P52732	-1.02	1.01	-0.45
TBC1 domain family member 10B	Q4KMP7	-1.05	0.21	2.07

Although the validation targets would be most likely selected from the list of 76 common proteins, the entire list of high confidence protein interaction partners for each NS5 protein was used for STRING and DAVID analysis to identify potential proteins and cellular pathways that were enriched in the respective FLAG pulldowns for the different NS5-FLAG proteins (Figures 35-43; Tables 9-17)). The identification of cellular pathways that were potentially influenced by the flaviviral NS5 proteins could also be used to guide the choice of validation targets towards biologically important NS5 binders by selecting proteins in pathways clearly heavily targeted by the NS5 proteins and pathways known to be targeted by NS5 from previous studies.

Figure 35:

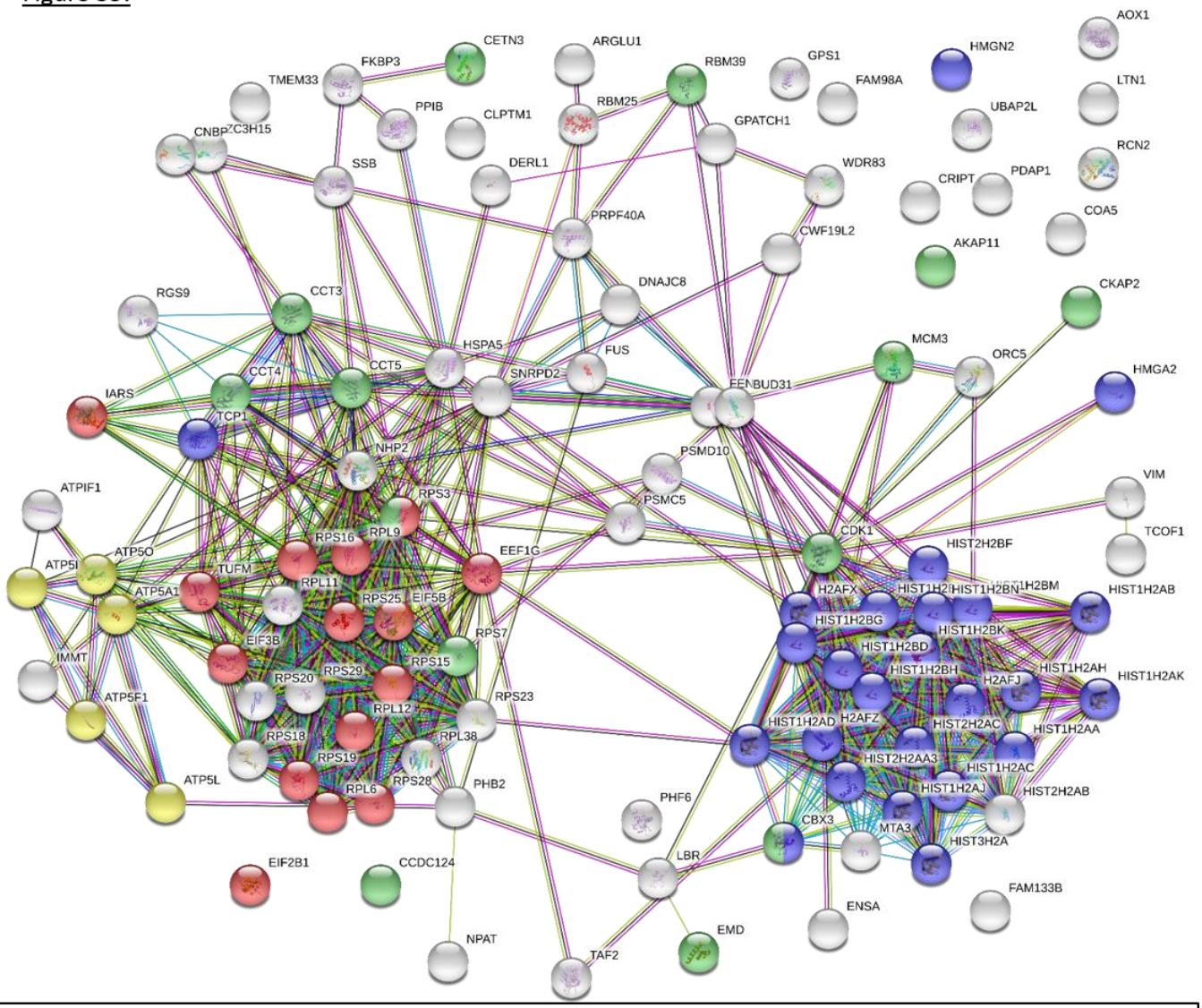

Figure 35: *STRING pathway analysis results on ZNS5-FLAG interactors in MaxQuant 0.01FDR* – High confidence interactors of ZNS5 generated from a ZNS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), histones/chromatin (blue), cytoskeleton (green), and mitochondrial inner membrane (yellow)

Figure 35 shows the STRING results for high confidence ZNS5 interactors in the 0.01FDR MaxQuant dataset with main functional annotation clusters from DAVID analysis of the same data highlighted. The top DAVID results and details from the STRING analysis are also presented in table 9 where it is clear that RNA/translation control is highly enriched as well as structural/cytoskeletal proteins.

Table 9: *DAVID/STRING functional annotation clustering for high confidence interactors of ZNS5 in the MaxQuant 0.01FDR dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	17.16	8.99e-14
Histone/chromatin	13.1	4.25e-16
Unfolded protein binding/microtubule	3.95	0.0127
Mitochondrial protein/inner membrane	3.36	0.0107
Actin	2.96	0.018

Figures 36 and 37 show the STRING analysis on the high-confidence proteins from MaxQuant 0.05FDR and PD for comparison.

Figure 36:

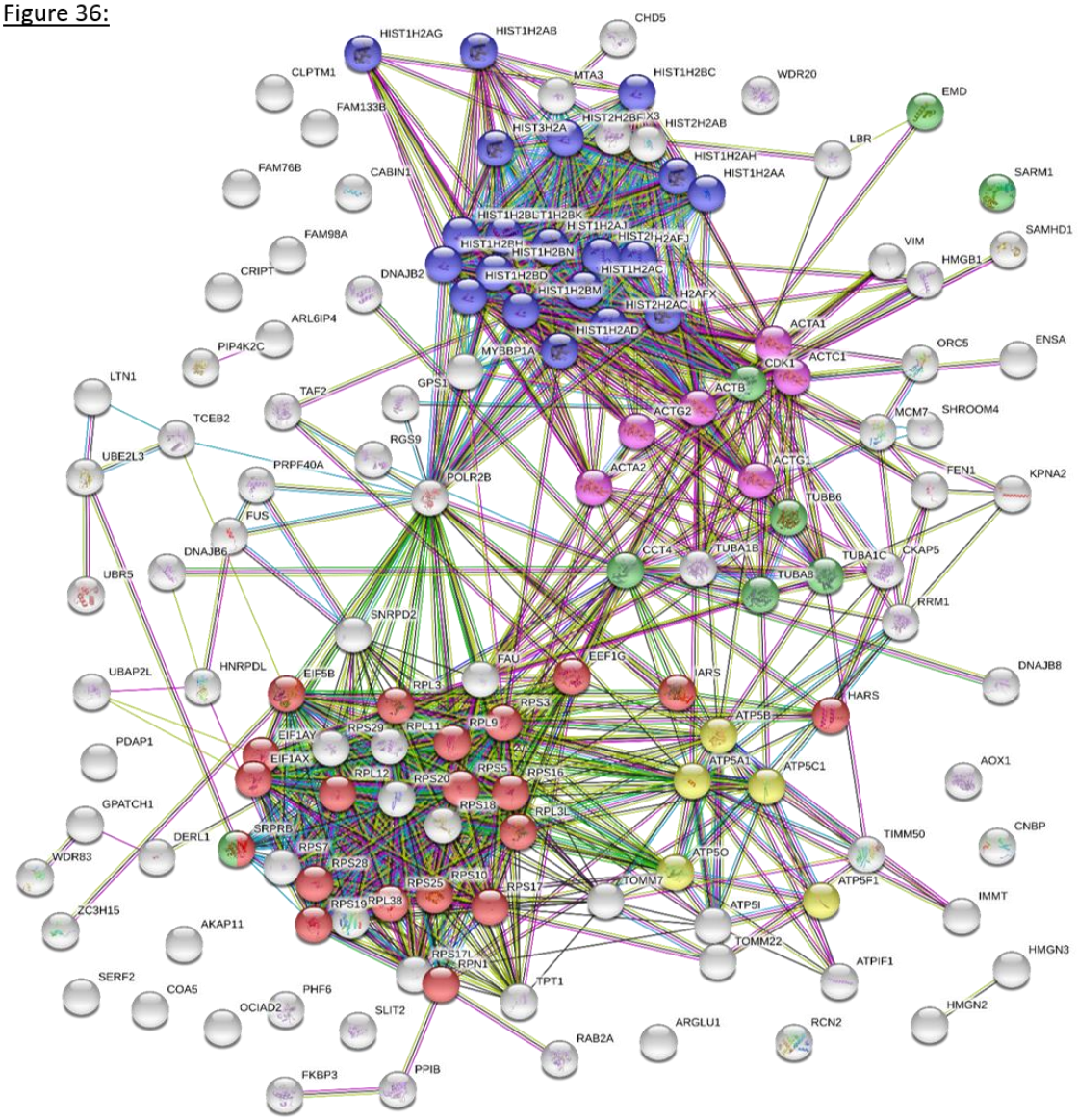

Figure 36: *STRING pathway analysis results on ZNS5-FLAG interactors in MaxQuant 0.05FDR* – High confidence interactors of ZNS5 generated from a ZNS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), histone/nucleosome (blue), microtubule proteins (green), mitochondria inner membrane/ATP synthesis (yellow), and actin (pink)

The proteins discovered in 0.05FDR MaxQuant show a lot of similarity in the STRING analysis (Figure 36) to the 0.01FDR with most of the same main clusters appearing. DAVID analysis (Table 10) also showed similar enrichment groups as the 0.01FDR MaxQuant data which are displayed in Figure 35 and Table 9.

Table 10: *DAVID/STRING functional annotation clustering for high confidence interactors of ZNS5 in the MaxQuant 0.05FDR dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	13.32	1.68e-11
Histone/nucleosome	12.94	4.1e-13
Actin	2.88	8.76e-07
Mitochondrial inner membrane/ATP synthesis	2.81	1.2e-06
Microtubule	2.35	0.0459

Figure 37:

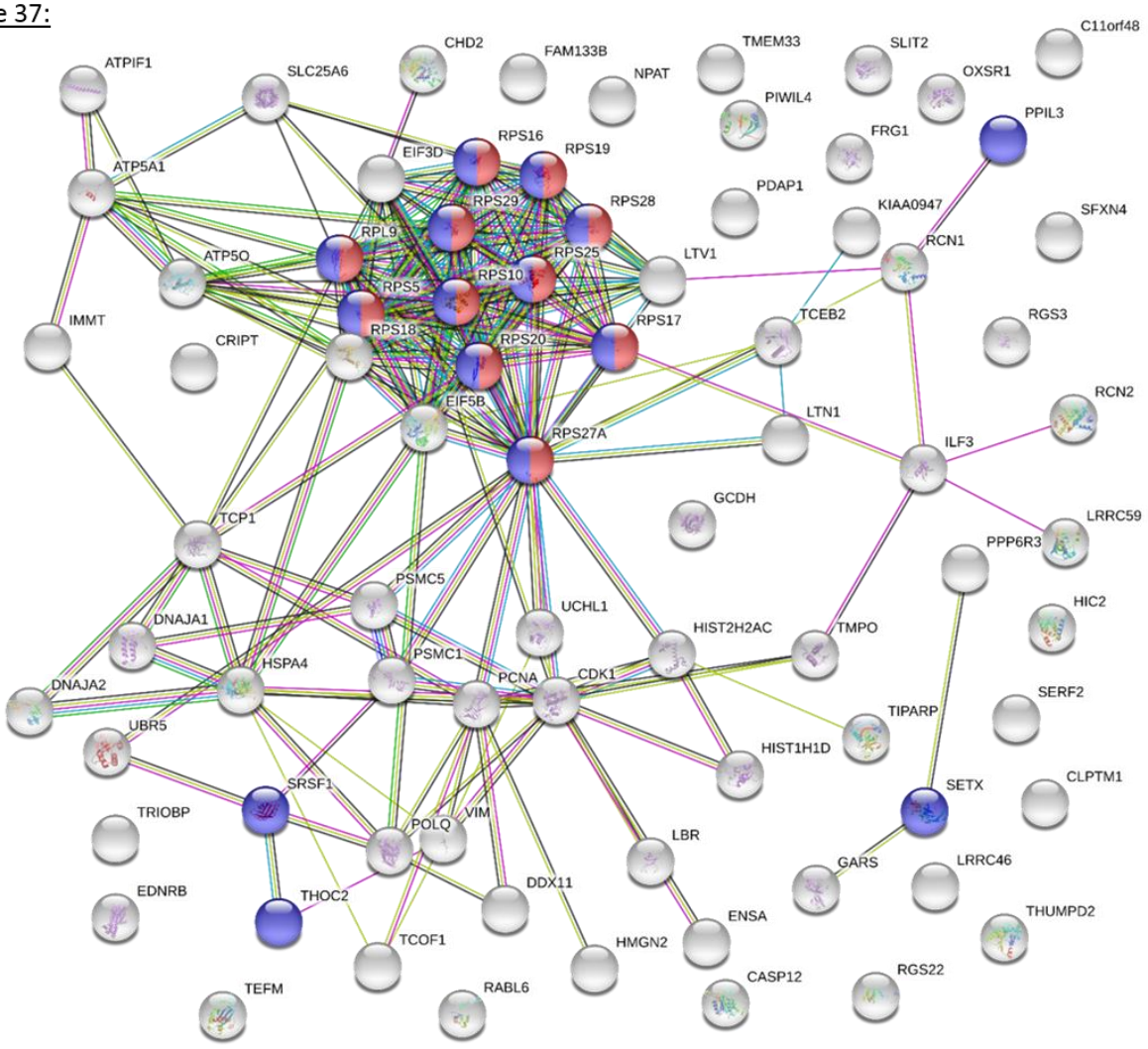

Figure 37: *STRING pathway analysis results on ZNS5-FLAG interactors in Proteome Discoverer* – High confidence interactors of ZNS5 generated from a ZNS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), and mRNA metabolic processes (blue)

Figure 37 shows PD's ZNS5 high-confidence protein STRING analysis. The results are far less clustered than that of the MaxQuant analyses and DAVID analysis only returned one high level enrichment which was the translation/ribosome proteins highlighted in Figure 37. The two following groups of ATP synthesis/mitochondria and mRNA splicing/processing are also displayed in Figure 37 as well as table 11 where the enrichment scores and false discovery rates show the large difference. Fewer high-confidence proteins identified by PD hits were found to form enriched clusters by STRING and DAVID analysis as many were proteins of unknown origin and were not identifiable by STRING or DAVID. They therefore did not appear on the STRING/DAVID analysis. Those that did map were far less clustered causing STRING to fail to present ATP synthesis/mitochondria as an enriched category to highlight and determine the false discovery rate of. Alternatively, it may have detected this enrichment but considered the false discovery rate to be >0.05.

Table 11: *DAVID/STRING functional annotation clustering for high confidence interactors of ZNS5 in the PD dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	9.73	1.78e-09
ATP synthesis/mitochondria	2.1	-
mRNA splicing/processing	1.29	9.98e-07

The D2NS5 STRING analysis is shown in Figures 38, 39, and 40 for 0.01FDR MaxQuant, 0.05FDR MaxQuant, and PD respectively. Once again, the differences between MaxQuant (Figures 38, 39) and PD (Figure 40) are clear. MaxQuant produces very clear defined clusters with greater enrichment and p values while PD analysis produced far fewer enrichment clusters. Tables 12-14 show this difference.

Figure 38:

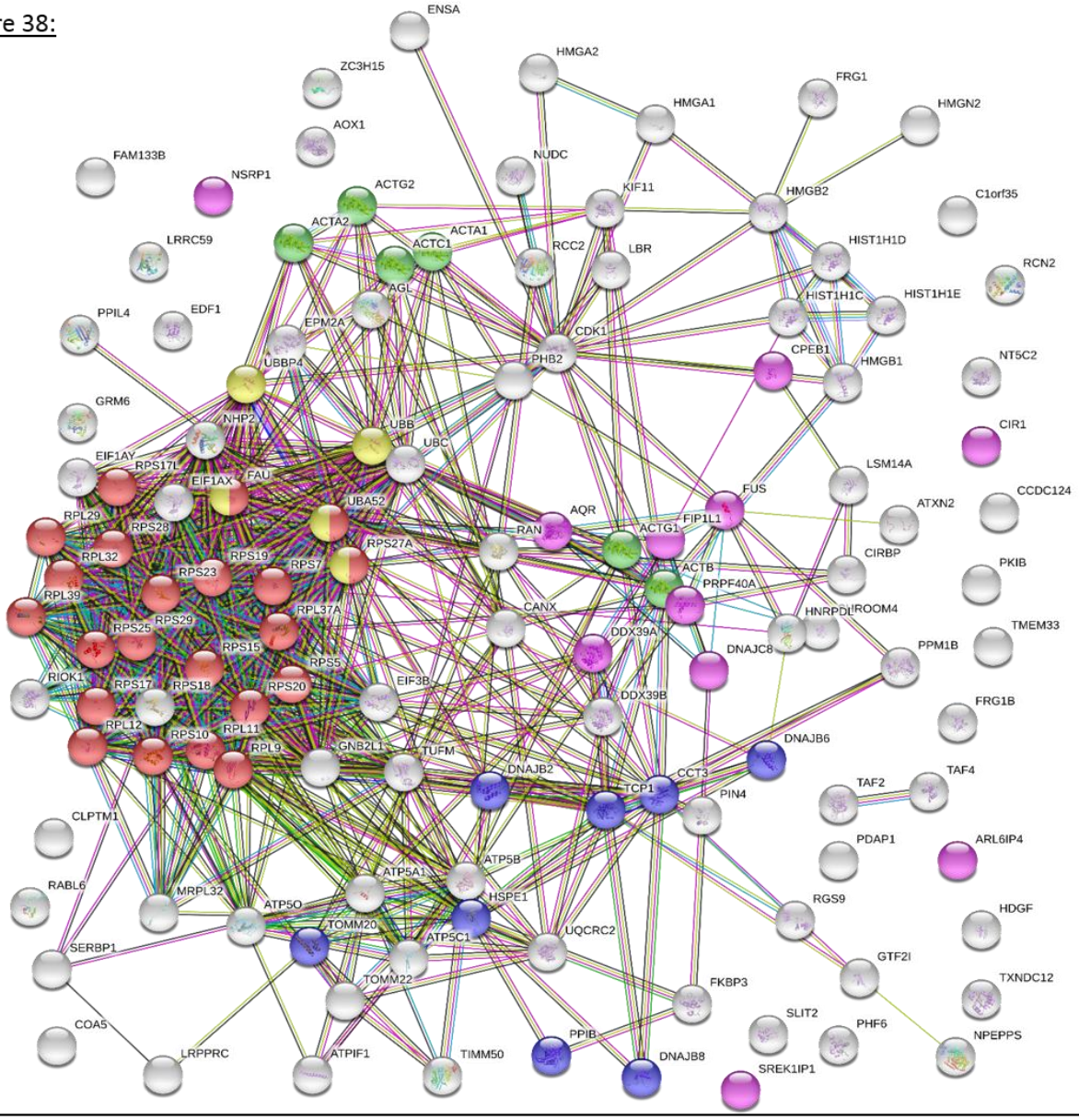

Figure 38: *STRING pathway analysis results on D2NS5-FLAG interactors in MaxQuant 0.01FDR* – High confidence interactors of D2NS5 generated from a D2NS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), Unfolded protein binding (blue), actin (green), and ubiquitin-related proteins (yellow), mRNA processing (pink), and mitochondria inner membrane/ATP synthesis (dark green).

D2NS5 enriched ubiquitin-related proteins in the FLAG pulldown (table 12); a large pool of proteins that involves ubiquitin E3 ligases, regulators of ubiquitination, and proteins involved in ubiquitin-dependent pathways such as DNA repair or protein degradation. It also shows enrichment for ribosome/translation which is expected in viral, and particularly flaviviral, proteins. mRNA processing is also expected with the known interaction between DENV-NS5 and the spliceosome.

Table 12: *DAVID/STRING functional annotation clustering for high confidence interactors of D2NS5 in the MaxQuant 0.01FDR dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	20.88	4.9e-24
Unfolded protein binding/chaperone	10.77	4.34e-06
Actin	2.81	1.05e-06
Ubiquitin-related proteins	2.7	2.47e-06
mRNA processing	2.65	0.00226
Mitochondria inner membrane/ATP synthesis	2.33	0.000648

The DAVID analysis of 0.05FDR MaxQuant showed the same enrichments as 0.01FDR MaxQuant though with the added enrichments of microtubule-related proteins (Figure 39)

Figure 39:

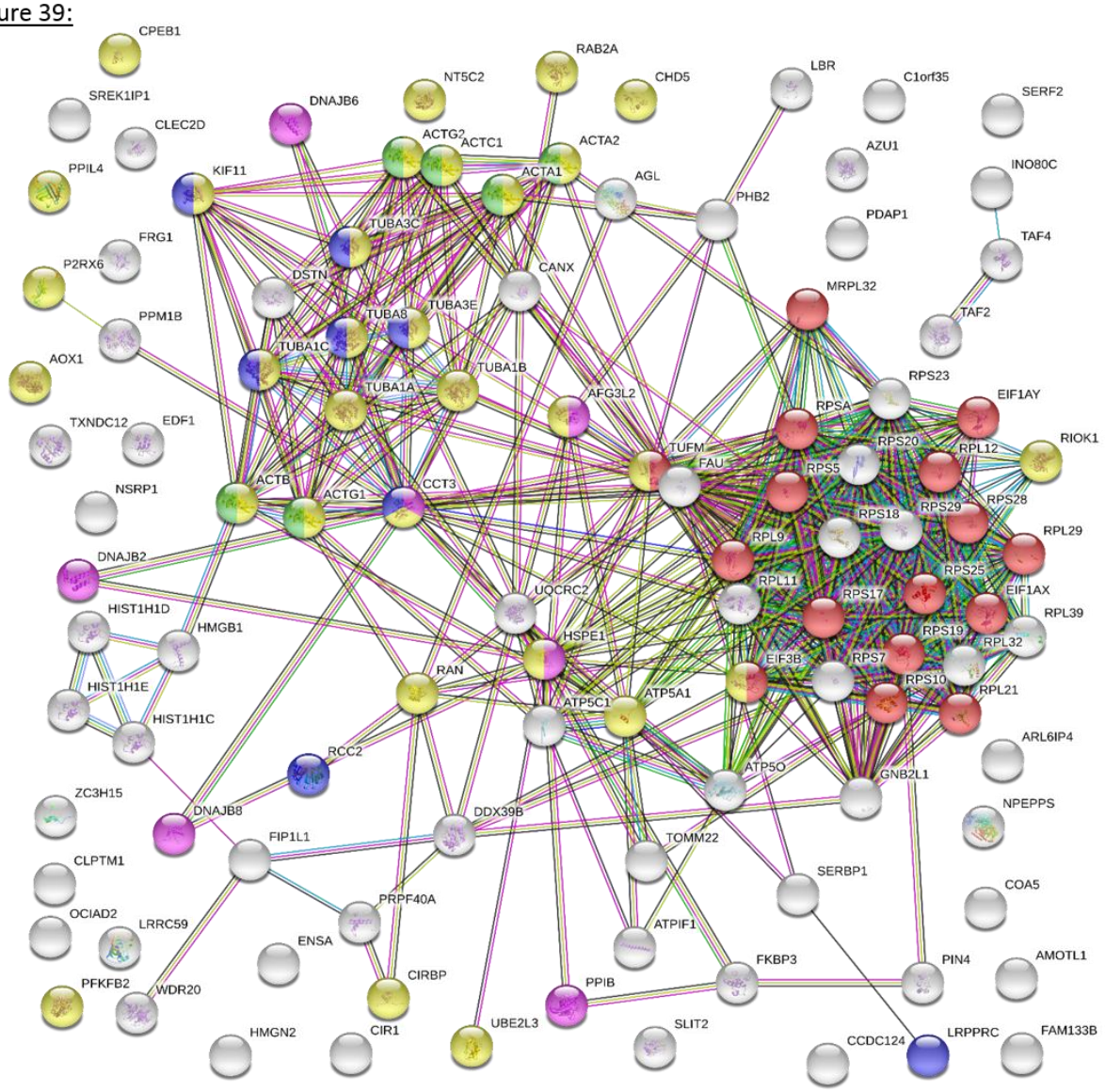

Figure 39: *STRING pathway analysis results on D2NS5-FLAG interactors in MaxQuant 0.05FDR* – High confidence interactors of D2NS5 generated from a D2NS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), microtubules (blue), actin (green), nucleotide binding (yellow), and unfolded protein binding/chaperone (pink)

Table 13: *DAVID/STRING functional annotation clustering for high confidence interactors of D2NS5 in the MaxQuant 0.05FDR dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	17.54	1.14e-19
Microtubule	4.96	0.0162
Actin	3.22	4.83e-07
ATP-/Nucleotide-binding	2.83	0.000146
Unfolded protein binding/chaperone	2.53	3.62e-05

PD, as with ZNS5, provided a much less defined set of proteins for D2NS5 and only showed three functional annotation clusters within DAVID (Figure 40).

Figure 40:

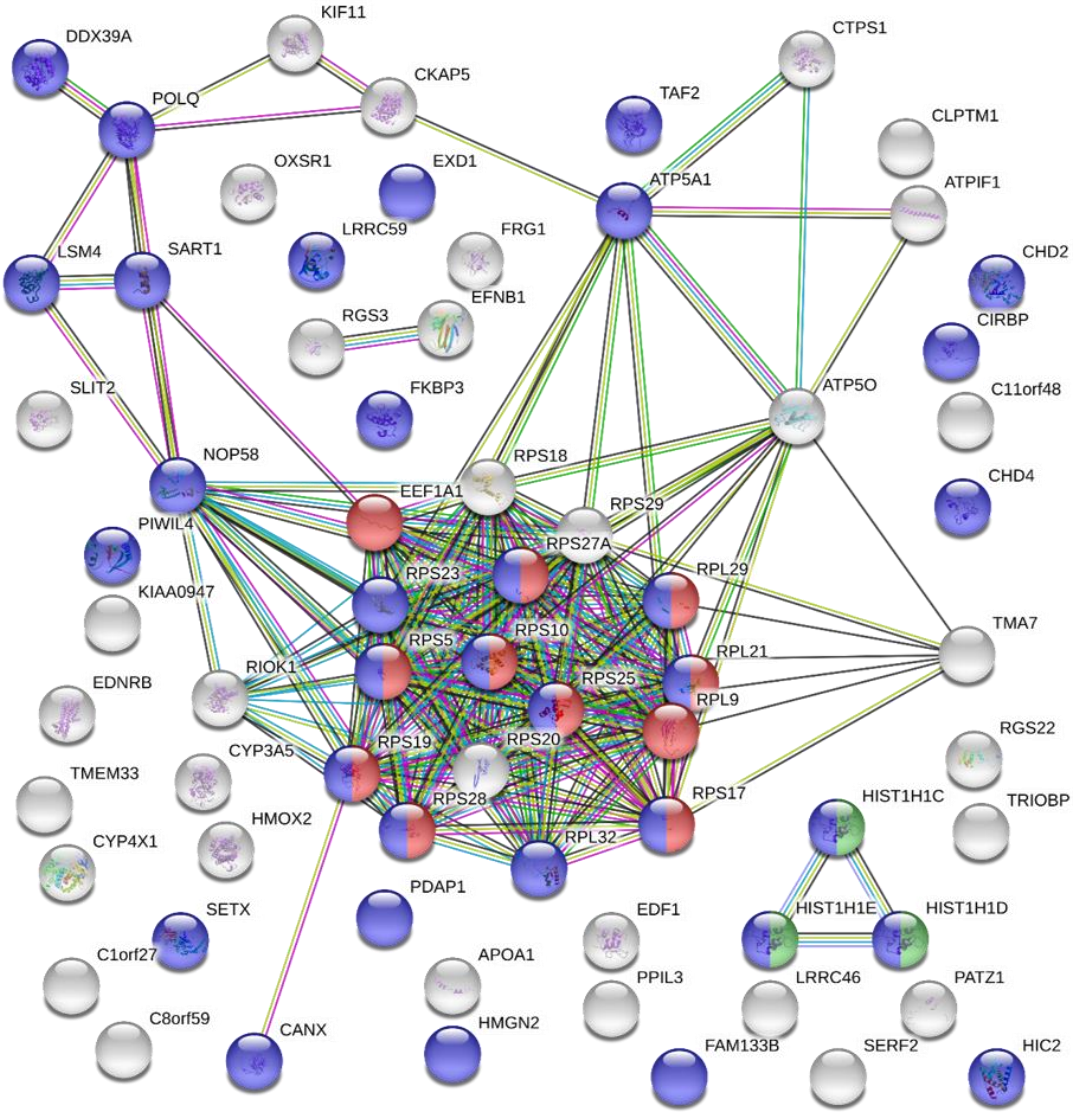

Figure 40: *STRING pathway analysis results on D2NS5-FLAG interactors in Proteome Discoverer* – High confidence interactors of D2NS5 generated from a D2NS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), nucleotide binding proteins (blue), and nucleosome proteins (green)

Table 14: *DAVID/STRING functional annotation clustering for high confidence interactors of D2NS5 in the PD dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	12.89	3.13e-06
Histone/nucleosome	2.32	0.00045
Helicase/ATP binding	1.46	0.0028

D4NS5 provided very clear protein clusters in all three datasets during STRING analysis. Figure 41 shows D4NS5 high-confidence protein interactors from the 0.01FDR analysis MaxQuant.

Figure 41:

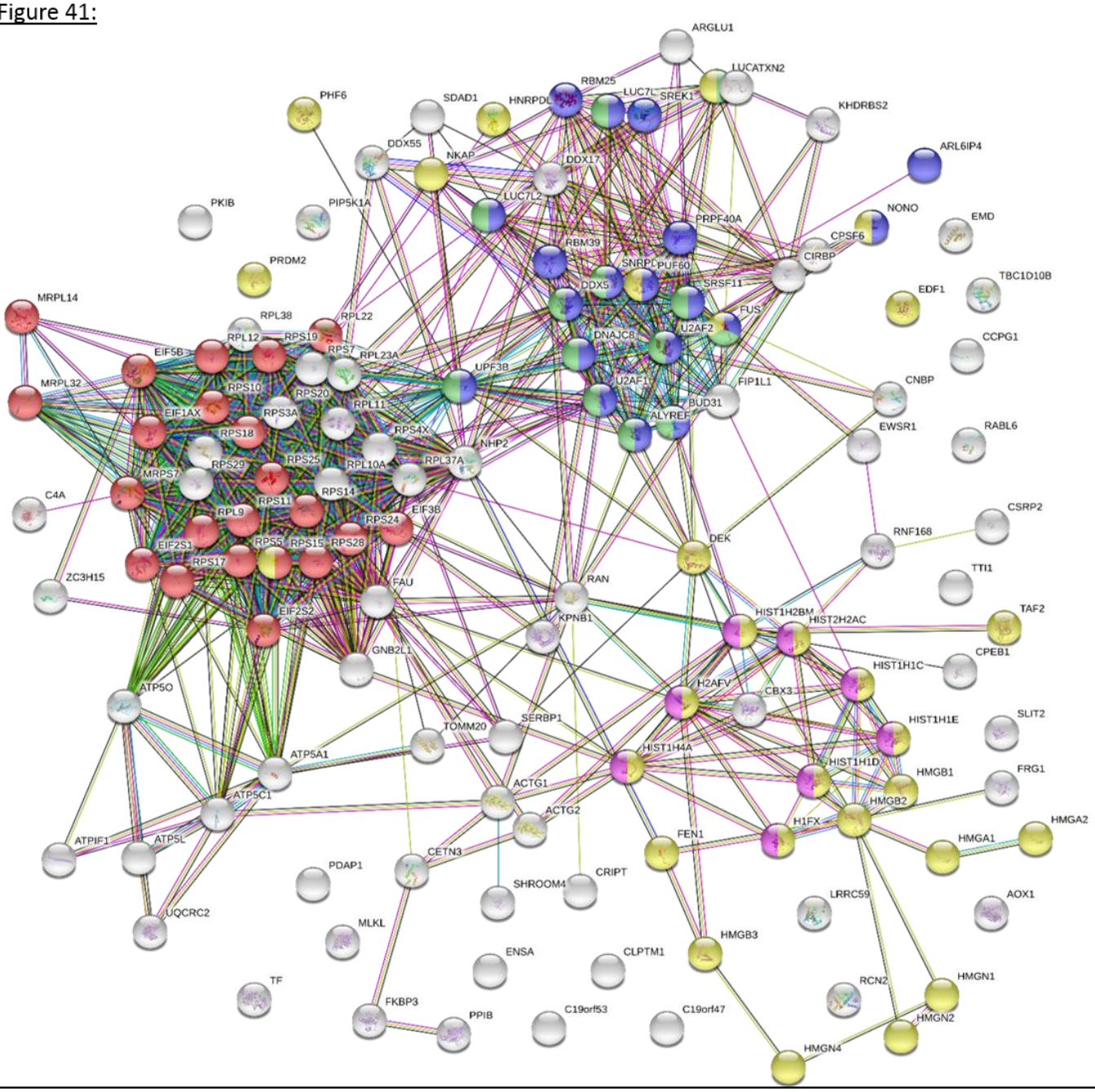

Figure 41: *STRING pathway analysis results on D4NS5-FLAG interactors in MaxQuant 0.01FDR* – High confidence interactors of D4NS5 generated from a D4NS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), RNA splicing (blue), mRNA splicing, via spliceosome (green), DNA binding (yellow), and nucleosome (pink)

The top DAVID defined enrichment clusters are highlighted in Figure 41 and bears a high level of similarity to D2NS5 and ZNS5. ‘Ribosome/translation’ is the most highly enriched term and has the lowest false discovery rate (Table 15) though ‘chromosome/DNA binding’ and ‘nucleosome/histone’ are enriched for D4NS5 and not for D2NS5 which is the more nuclear-localised of the two.

Table 15: *DAVID/STRING functional annotation clustering for high confidence interactors of D4NS5 in the MaxQuant 0.01FDR dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	23.44	4.26e-14
mRNA processing/binding/splicing	9.05	5.46e-11
Spliceosome/mRNA export	5.56	1.49e-07
Chromosome/DNA binding	3.47	0.0236
Nucleosome/histone	3.45	2.94e-06

Figure 42 shows the STRING analysis for 0.05 MaxQuant D4NS5 protein interactors where the five main clusters from the 0.01FDR MaxQuant STRING analysis can still be easily seen.

Figure 42:

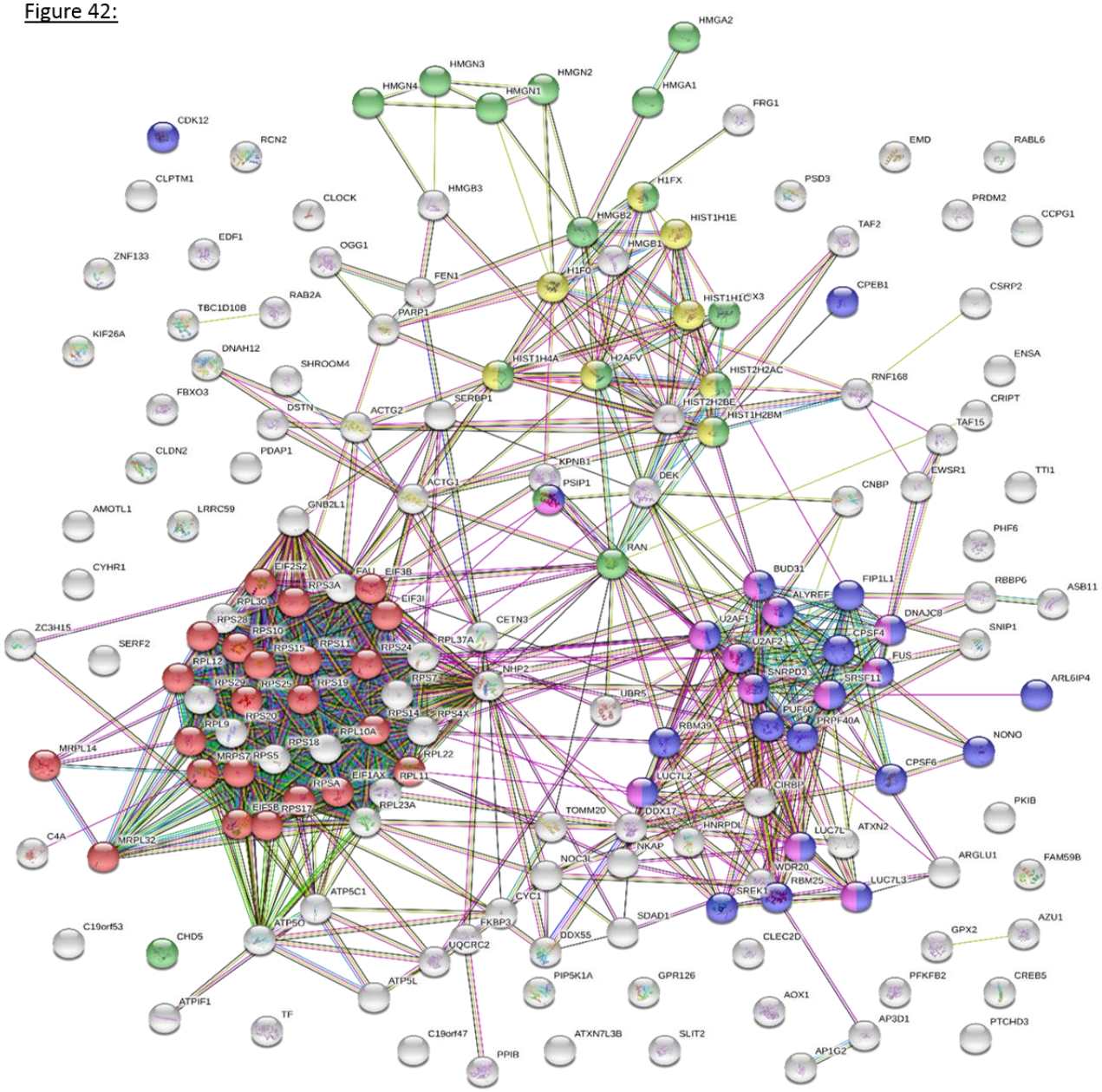

Figure 42: *STRING pathway analysis results on D4NS5-FLAG interactors in MaxQuant 0.05FDR* – High confidence interactors of D4NS5 generated from a D4NS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), mRNA processing (blue), chromatin (green), nucleosome (yellow), and mRNA splicing, via spliceosome (pink)

Table 16: *DAVID/STRING functional annotation clustering for high confidence interactors of D4NS5 in the MaxQuant 0.05FDR dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	21.91	6.63e-13
mRNA processing/splicing/RNA binding	7.53	2.35e-12
Chromatin	4.53	4.34e-06
Nucleosome/chromosome	4.12	1.63e-05
Spliceosome	3.53	2.13e-05

Finally, PD's D4NS5 protein interactor list was processed through STRING and DAVID and, although less defined compared to the MaxQuant datasets, still showed a clear clustering of ribosomal/translation initiation proteins (Figure 43). The remaining categories identified as highly enriched by DAVID are less clearly defined in STRING but were; mRNA processing, cell-cell adhesion, chromatin, and RNA transport proteins.

Figure 43:

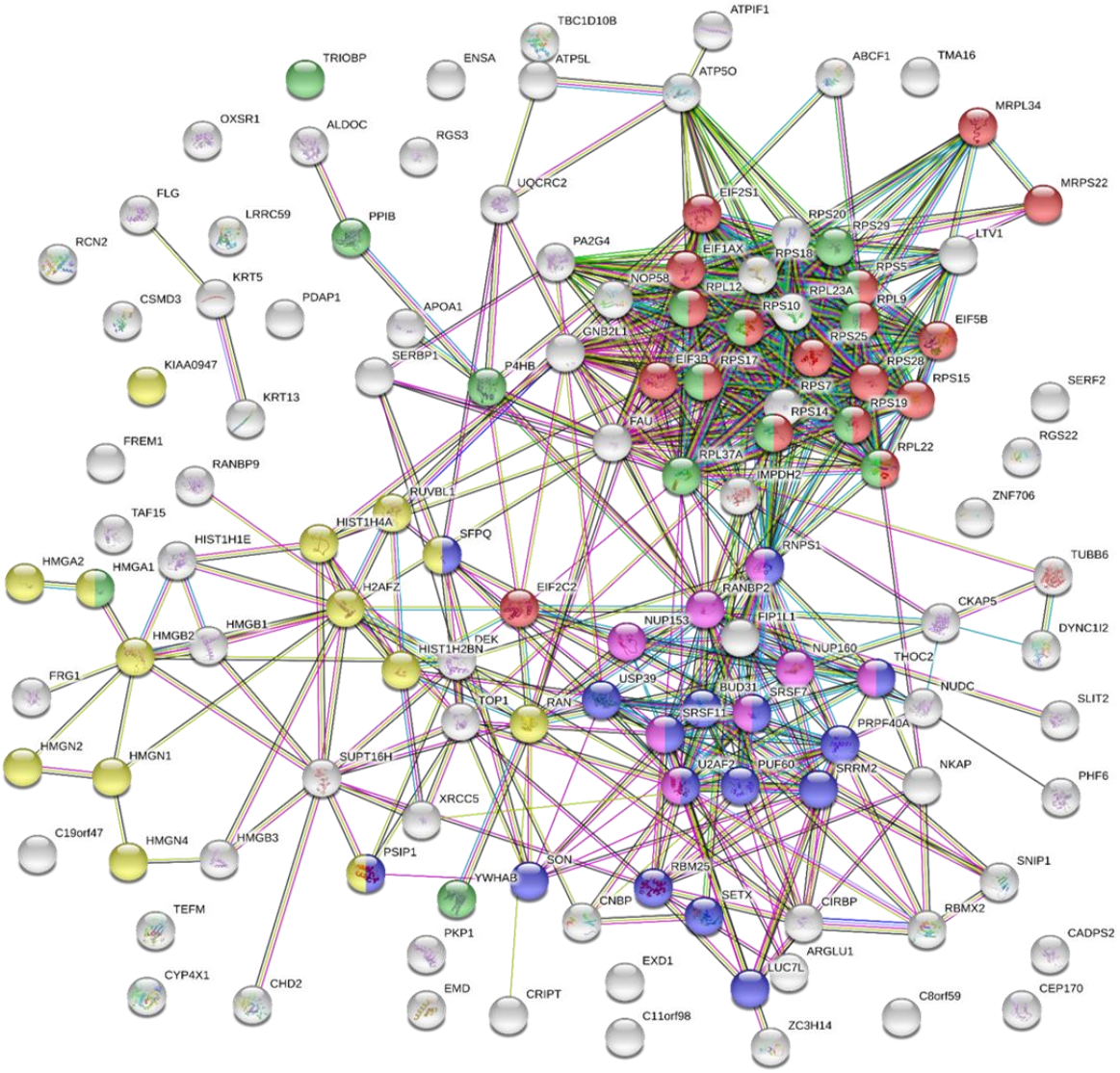

Figure 43: *STRING pathway analysis results on D4NS5-FLAG interactors in Proteome Discoverer* – High confidence interactors of D4NS5 generated from a D4NS5-FLAG pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; translation proteins (red), RNA splicing (blue), focal adhesion (green), chromatin (yellow), and mRNA transport (pink)

Table 17: *DAVID/STRING functional annotation clustering for high confidence interactors of D4NS5 in the PD dataset*

Functional annotation cluster	EASE score	STRING false discovery rate
Ribosome/translation	15.08	3.66e-10
mRNA splicing/processing	5.74	2.8e-08
Cell-cell adhesion	5.4	3.19e-07
Chromatin	3.73	3.84e-06
RNA transport	2.61	0.000111

Beyond the highest enriched categories, there were also a number of borderline enriched categories within DAVID that did not appear clearly within STRING and so an exhaustive list of DAVID functional annotation clusters for each NS5 protein's high confidence interactor list from MaxQuant 0.01FDR has been provided in table 18. The categories are organised from highest enrichment to the lowest for each dataset from top to bottom.

Table 18: *Full list of DAVID functional annotation clusters for the high confidence interactors of each NS5 protein*

D2NS5	EASE score	D4NS5	EASE score	ZNS5	EASE score
Ribosome/ translation	20.88	Ribosome/ translation	23.44	Ribosome/ translation	17.16
Unfolded protein binding/ chaperone	10.77	mRNA processing/ binding/ splicing	9.05	Histone/ chromatin	13.1
Actin	2.81	Spliceosome/ mRNA export	5.56	Unfolded protein binding/ microtubule	3.95
Ubiquitin-related proteins	2.7	Chromosome/ DNA binding	3.47	Mitochondrial protein/ inner membrane	3.36
mRNA processing	2.65	Nucleosome/ histone	3.45	Actin	2.96

Mitochondria inner membrane/ ATP synthesis	2.33	Chromatin	3.43	Protein biosynthesis/ RNA transport	2.78
Rotamase/ Isomerase	2.27	RNA transport	2.61	Nucleotide binding/ ATP binding	2.32
Histone/ chromosome	1.96	Actin	2.21	Unfolded protein binding/ DnaJ	2.03
Mitosis/ cell division	1.79	DNA binding/ DNA bending	1.81	Spliceosome	1.3
Spliceosome	1.56	ATP synthesis	1.38	Cell cycle/ division	1.3
Cell-cell adhesion	1.28	DNA damage/ DNA repair/base excision repair	1.37	DNA damage/ DNA repair	1.3
		Cell-cell adhesion	1.31		

DNA damage is one of the most notable hallmarks of ZIKV infection with the development of pyknotic nuclei a recognised feature of ZIKV-infected cells (54) so the appearance of ‘DNA damage/DNA repair’ as a functional annotation cluster was unsurprising. However, ‘DNA damage/DNA repair’ was also seen in D4NS5 at a slightly higher level of enrichment than in ZNS5 high confidence interactors. So far there have been no reports of DENV-NS5 causing DNA damage like ZNS5. Closer analysis of the proteins involved in the ZNS5/D4NS5 ‘DNA damage/DNA repair’ cluster has been shown in tables 19, 20, and 21 which revealed numerous of D4NS5’s DNA damage/DNA repair proteins to have additional roles in other pathways.

Table 19: *DNA damage/DNA repair cluster proteins in both ZNS5 and D4NS5*

Protein name	UniProt accession	Function
Histone H2AX	P16104	Histone protein - involved in regulation of numerous DNA processes like repair, transcription, and replication
Flap endonuclease 1	P39749	Cuts 5' RNA flap from okazaki fragments as well as performing endonuclease and exonuclease activities within base excision repair and in the cases of nicked DNA or ds breaks as part of the repair process
High mobility group protein B1	P09429	Nucleosomal member of a group of proteins that regulate numerous DNA functions including replication, repair, apoptosis, nucleosome rearrangement, and transcription as well as a large number of intracellular processes. It can also function as a DAMP, mediating the response to inflammation, infection, and damage when secreted extracellularly

Table 20: *DNA damage/DNA repair cluster proteins in ZNS5 only*

Protein name	UniProt accession	Function
Cyclin-dependent kinase 1	P11440	Cell cycle regulation protein (G1→ S phase) with large number of intracellular targets for phosphorylation including HSP70; a DNA protective protein
DNA replication licensing factor MCM7	P33993	DNA helicase responsible for cell cycle progression through S phase - detects UV-related DNA damage
40S ribosomal protein S3	P23396	Contains endonuclease activity used in DNA repair - has a high affinity for ROS-damaged guanine within DNA

E3 ubiquitin-protein ligase UBR5	O95071	Ubiquitin ligase - activates CDK9 in DNA maturation, regulates TopBP1 and RNF168 (both involved in DNA repair), and also theorised to have a role in cell cycle progression

Table 21: *DNA damage/DNA repair cluster proteins in D4NS5 only*

Protein name	UniProt accession	Function
N-glycosylase/DNA lyase	O15527	DNA incision protein with a high affinity for ROS-damage guanine within DNA
E3 ubiquitin-protein ligase RBBP6	Q7Z6E9	Ubiquitin ligase - encourages the degradation of YBX1 as well as promoting cell growth and apoptosis via the degradation of p53
Circadian locomoter output cycles protein kaput	O15516	TF involved in the regulation of the circadian rhythm - repressed by p53 and NONO
High mobility group protein HMG-I/Y	P17096	Regulates expression in dsDNA regions with high AT concentrations - could be involved in nucleosome placement
High mobility group protein HMGI-C	P52926	Regulator of transcription and plays a role in cell cycle progression via interactions with Cyclin-A2
High mobility group protein B2	P26583	Involved in histone remodelling and transcription regulation within the nucleus and is reported to be involved in DNA ds break repair ligation
Non-POU domain-containing octamer-binding protein	Q15233	Repressor of CLOCK and a target of CDK1 phosphorylation. Interacts with U5 snRNA and could be involved in non-homologous end joining in ds breaks

Poly[ADP-ribose] polymerase 1	P09874	Utilised in base excision repair of damaged DNA and acts as a transcription regulator for specific genes
E3 ubiquitin-protein ligase RNF168	Q8IYW5	Ubiquitin ligase - encourages recruitment of DNA repair proteins to sites of DNA damage particularly are sites of ds breaks

The differences between 0.01FDR and 0.05FDR MaxQuant are as expected with the 0.05FDR dataset having a greater number of proteins detected and slightly fewer high confidence proteins as well as higher p values in STRING analysis and lower enrichment values in DAVID analysis but still retaining a high degree of similarity in the functional annotation clusters. The differences between PD and MaxQuant however are much more noticeable. Table 7 shows how few common proteins were found between the two programmes when compared to the comparison of the two FDRs within the same programme. PD also provided fewer DAVID categories in general and less clear STRING clusters as shown in Figures 37, 40, and 43, partly due to a tendency to assign detected peptides to poorly reported proteins that were not on the STRING/DAVID database. It was decided because of this that the Streptavidin pulldown proteomic data would initially be analysed with MaxQuant set at an FDR of 0.01 with the possibility to reanalyse the data afterwards with PD or a higher FDR if desired.

3.6.3: Streptavidin pulldown bioinformatics

The streptavidin pulldown samples described in Section 3.5.2 were processed using TMT-MS/MS and the resulting data files analysed by MaxQuant at an FDR of 0.01 using the same protein search list as the FLAG pulldown samples with the addition of the ZNS5-BirA sequence. The resulting data was analysed first in Perseus (as in 3.6.2) where proteins detected by only a single site were removed from the sample set. A log2 transformation was performed on the raw values which were then normalised with a median subtraction to provide a normal distribution around the value of log2 value of '0'. Histograms were generated using Perseus to show the normal distributions and are shown in Figure 44.

Figure 44:

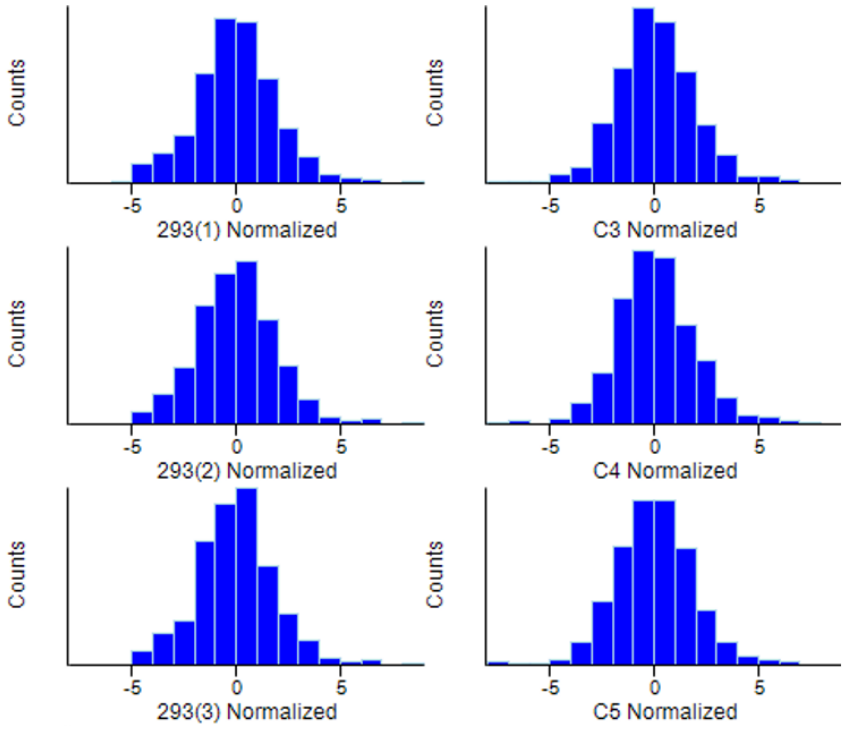

Figure 44: *Histograms showing the distribution of protein concentrations of each Streptavidin pulldown sample from the 0.01 FDR MaxQuant-processed TMT-MS/MS data* – Proteomic data from the MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through MaxQuant at a FDR of 0.01 before being processed in Perseus. A log2 transformation was performed on the protein counts and the data was then normalised using median subtraction to give the shown histograms

Scatterplots were produced with Pearson's Correlation values to show the correlation of the replicates with each other. As shown in Figure 45, the replicates for the 293-Flp cells were highly reproducible and reliable as were the datasets using the three independently produced ZNS5-BirA cell lines (C3, C4, C5).

Figure 45:

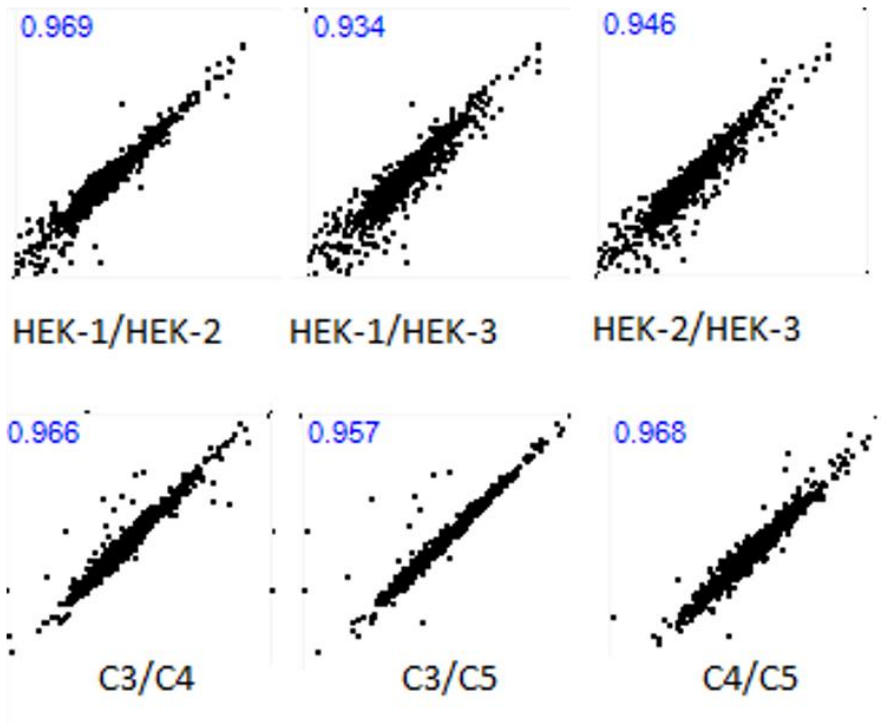

Figure 45: *Scatterplot graphs showing the correlation between Streptavidin pulldown samples from the 0.01FDR MaxQuant-processed TMT-MS/MS data* - Proteomic data from the MS/MS of NS5 interactors and HEK293 FLP-In™ control cells was passed through MaxQuant at a FDR of 0.01 before being processed in Perseus where scatterplots were generated for replicates of each cell line along with the Pearson's correlation value in the top left of each individual scatterplot.

The pull down data produced using the C5 cell line (Figure 44 and 45 as well as raw values) suggested that the C5 pulldown was comparable to C3 and C4 and it was therefore included in the final analysis. As with the MaxQuant data in Section 3.6.2, the dataset for the streptavidin pulldown was analysed using a two sample t-test to generate a t-test difference value and a –log10 p value for each protein detected. This data was then exported to Microsoft Office Excel to for further analysis and is provided on the supplementary CD in the appendix (filename; 'BirA').

The dataset was analysed using the same criteria as the FLAG pulldown data; a log2 t-test difference of >0.58 and a –log10 p value of >1.3 being the lower boundary for selection. No proteins with a log2 t-test difference between 0.50-0.58 were added to this dataset and all proteins detected within the dataset had >1 peptide detected and a –log10 p value >1.3. Table 22 shows a summary of the processed dataset.

Table 22: *Quantity of streptavidin pulldown raw hits and high confidence proteins*

No. of proteins detected	No. of proteins log2 t-test>0.58	No. of proteins –log10 p value >1.3	No. of final high confidence proteins
812	275	553	247

As shown in Table 22, there was a larger number of proteins detected compared to the FLAG pulldown and a higher number of high confidence proteins. The full list is shown in Table 23.

Table 23: *High confidence protein list from streptavidin pulldown on ZNS5-BirA cell lines*

Protein name	Uniprot accession	log2 t-test difference	-log10 p value
Small acidic protein	O00193	4.629092	2.530277
TOX high mobility group box family member 4	O94842	4.228101	1.784686
General transcription factor IIE subunit 1	P29083	4.091256	1.762518
cDNA, FLJ94590, highly similar to Homo sapiens craniofacial development protein 1 (CFDP1), mRNA	B2R9W9	3.968232	2.058638
Zinc finger CCCH domain-containing protein 18	E7ERS3	3.651489	5.517842
Treacle protein	E7ETY2	3.520173	2.741643
ELL-associated factor 1	Q96JC9	3.496523	1.979384
PEST proteolytic signal-containing nuclear protein	Q8WW12	3.379783	2.494137
RNA polymerase II elongation factor ELL2	O00472	3.310556	1.974359
Transcriptional regulator Kaiso	Q86T24	3.297905	2.358113
DDB1- and CUL4-associated factor 5	Q96JK2	3.292729	2.43229
YEATS domain-containing protein 2	Q9ULM3	3.199635	2.608447
UPF0428 protein CXorf56	Q9H5V9	3.163606	2.099316
WW domain-binding protein 11	Q9Y2W2	3.101506	2.95131
Treacle protein	Q13428	3.046453	2.12005
Tuftelin-interacting protein 11	Q8N523	3.039954	2.190444
ZNF507 protein	B9EGE7	3.029881	1.889423
Spliceosome-associated protein CWC15 homolog	Q9P013	2.979059	1.634118

Zinc finger CCCH domain-containing protein 4	Q9UPT8	2.972613	4.282398
RNA cytidine acetyltransferase	Q9H0A0	2.959367	2.374478
Targeting protein for Xklp2	Q9ULW0	2.825635	2.213261
Nucleoporin 153kDa, isoform CRA_a	A0A024QZW7	2.825548	2.372861
Phosphorylated adapter RNA export protein	Q9H814	2.811495	2.860488
Transcription initiation factor IIE subunit beta	P29084	2.793195	2.159655
Transcription initiation factor TFIID subunit 9	Q16594	2.758743	2.242561
Histone-lysine N-methyltransferase SETD2	Q9BYW2	2.731504	2.103004
cDNA FLJ76716, highly similar to Homo sapiens WD repeat domain 70 (WDR70), mRNA	A8K564	2.638719	3.070081
cDNA FLJ30322 fis, clone BRACE2006703, highly similar to Surfeit locus protein 6	B3KNS8	2.633578	3.093013
Negative elongation factor E	H9ZYJ1	2.632782	3.620597
Bcl-2-associated transcription factor 1	E9PK91	2.615508	3.448974
NMDA receptor regulated 2, isoform CRA_b	A0A024R5V9	2.60136	2.110984
HIV Tat-specific factor 1	O43719	2.591445	2.664477
Coilin	P38432	2.537444	2.336884
Calcium homeostasis endoplasmic reticulum protein	J3QK89	2.50319	1.950664
SAP30-binding protein (Fragment)	J3QQJ0	2.478776	3.572543
RAD21 homolog (S. pombe), isoform CRA_a	A0A024R9J0	2.42804	3.958031
SURP and G-patch domain-containing protein 2	M0R2Z9	2.426897	2.66168
Zinc finger protein 447, isoform CRA_a	A0A024R4T0	2.414669	2.678586
cDNA FLJ58676, highly similar to Protein AF-9	B7Z4N5	2.408514	1.684084
Little elongation complex subunit 1	Q9Y2F5	2.395516	2.288308
Nucleoprotein TPR	P12270	2.319734	2.367651

cDNA FLJ76043, highly similar to Homo sapiens cyclin T1 (CCNT1), mRNA	A8K4M5	2.228825	3.311964
Paired amphipathic helix protein Sin3b	O75182	2.223139	2.922717
Nuclear pore complex protein Nup88	Q99567	2.217167	2.411873
Pre-mRNA-processing factor 17	O60508	2.210404	2.307844
Zinc finger MYM-type protein 1	Q5SVZ6	2.2075	2.111967
Ubiquitin carboxyl-terminal hydrolase 28	Q96RU2	2.198953	4.250165
Thyroid hormone receptor-associated protein 3	Q9Y2W1	2.17513	2.474572
Uncharacterized protein C19orf43	Q9BQ61	2.167783	2.599537
Cyclin K, isoform CRA_c	A0A024R6K1	2.166432	2.083365
Ran-binding protein 3	Q9H6Z4	2.156872	2.990789
E3 SUMO-protein ligase RanBP2	P49792	2.156651	2.263373
Kinesin-like protein KIF23	Q02241	2.153106	3.438644
Segment polarity protein dishevelled homolog DVL-3	Q92997	2.15095	1.981496
Filamin-A	P21333	2.138541	2.039296
Bromodomain-containing protein 4	O60885	2.137367	2.94827
Thymopoietin, isoform CRA_c	A0A024RBE7	2.110509	2.054822
Bcl-2-associated transcription factor 1	Q9NYF8	2.030242	2.201267
Mediator of RNA polymerase II transcription subunit 1	Q15648	2.014831	2.969941
U4/U6.U5 tri-snRNP-associated protein 1	O43290	2.004564	3.821661
Serine/threonine-protein phosphatase 4 regulatory subunit 3A	Q6IN85	1.992283	2.549191
RNA-binding protein 27	Q9P2N5	1.973138	2.773509
Chromodomain-helicase-DNA-binding protein 4	A0A0C4DGG9	1.965321	2.581496
PERQ amino acid-rich with GYF domain-containing protein 2	I1E4Y6	1.958627	2.471605
General transcription factor 3C polypeptide 4	Q9UKN8	1.939162	2.246829
Myb/SANT-like DNA-binding domain-containing protein 2	Q6P1R3	1.930214	1.623364
cDNA, FLJ95596, highly similar to Homo sapiens activity-dependent neuroprotector (ADNP), mRNA	B2RBM8	1.920433	1.30159

Nuclear factor 1	B1AKN8	1.904528	3.309829
UPF0690 protein C1orf52	Q8N6N3	1.896129	4.137287
Nucleolar and coiled-body phosphoprotein 1 (Fragment)	A0A0A0MRM9	1.892745	4.360396
Probable ATP-dependent RNA helicase DDX46	A0A0C4DG89	1.885934	2.529869
Splicing factor 3B subunit 1	O75533	1.880557	2.512787
RNA-binding protein 33	Q96EV2	1.872197	2.445714
Zinc finger DBF-type containing 2 vesion 1	N0DVB2	1.86688	2.158866
RNA-binding protein 25	P49756	1.864361	3.9205
BCL6 corepressor-cyclin B3 fusion protein	H9A532	1.861885	3.10134
cDNA FLJ77796, highly similar to Homo sapiens evolutionarily conserved G-patch domain containing protein mRNA	A8K8Y8	1.858805	2.206902
RNA-binding protein 26	A0A087X0H9	1.846651	3.267271
cDNA FLJ75279, highly similar to Homo sapiens elongation factor RNA polymerase II (ELL), mRNA	A8KAP0	1.831473	2.02775
Nuclear autoantigenic sperm protein	P49321	1.830069	3.631122
Negative elongation factor A	A0A0C4DFX9	1.827671	1.986882
PCM1 protein	B9EIS5	1.821753	2.070436
Cleavage and polyadenylation specificity factor subunit 6	Q16630	1.814633	4.639645
RNA binding motif protein 17, isoform CRA_a	Q5W009	1.778106	2.027115
Nuclear mitotic apparatus protein 1, isoform CRA_a	A0A024R5M9	1.758977	2.556108
Antigen KI-67	P46013	1.755822	2.061346
Splicing factor 4	A5PLN4	1.753743	2.736034
CASP8-associated protein 2	Q9UKL3	1.74339	1.733813
Protein PML	P29590	1.729818	3.597707
Chromodomain-helicase-DNA-binding protein 8	Q9HCK8	1.708258	2.722994
Zinc finger protein 318	Q5VUA4	1.705546	2.419869
Nibrin	A6H8Y5	1.697521	1.824391
Wings apart-like protein homolog	Q7Z5K2	1.67554	2.817996
Nuclear pore complex protein Nup50	Q9UKX7	1.669251	2.456212

RNA polymerase-associated protein LEO1	Q8WVC0	1.665639	2.610585
Zinc finger protein 609	O15014	1.661363	1.789352
Chromodomain helicase DNA binding protein 7, isoform CRA_a	A0A024R7V7	1.658542	2.562533
Serine/arginine-rich splicing factor 11	Q05519	1.650664	1.585411
Segment polarity protein dishevelled homolog DVL-2	O14641	1.648903	1.8389
Adapter molecule crk	P46108	1.648269	4.016389
Adenosine deaminase, RNA-specific isoform ADAR-a variant (Fragment)	Q59EC0	1.63492	2.715724
Nuclear cap-binding protein subunit 3	Q53F19	1.624838	1.631089
WD repeat-containing protein 36	Q8NI36	1.623085	1.507234
U2 small nuclear ribonucleoprotein A'	P09661	1.622078	3.638087
Nucleoporin 214kDa, isoform CRA_b	A0A024R8B6	1.608553	2.837137
RNA binding motif protein 10 isoform 1 (Fragment)	A0A0S2Z4X1	1.594113	5.35855
ATPase WRNIP1	Q96S55	1.593846	2.036543
Ankyrin repeat domain-containing protein 11	Q6UB99	1.579614	2.611768
Apoptotic chromatin condensation inducer in the nucleus	Q9UKV3	1.576627	2.706251
cDNA FLJ76962, highly similar to Homo sapiens nucleolar protein 5A (56kDa with KKE/D repeat) (NOL5A), mRNA	A8K9K6	1.571525	2.323066
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1	Q9H4L7	1.564526	2.151807
cDNA FLJ54492, highly similar to Eukaryotic translation initiation factor 4B	B4DRM3	1.563131	3.615853
cDNA, FLJ92684, highly similar to Homo sapiens IK cytokine, down-regulator of HLA II (IK), mRNA	B2R5Y4	1.56238	2.272685
60S ribosomal protein L10	F8W7C6	1.554221	2.394062
Paired amphipathic helix protein Sin3a	Q96ST3	1.544627	1.636043
EF-hand domain-containing protein D2	Q96C19	1.543405	1.754067

cDNA, FLJ93335, highly similar to Homo sapiens PRP3 pre-mRNA processing factor 3 homolog (yeast) (PRPF3), mRNA	B2R791	1.540373	2.276223
Importin subunit alpha-3	O00629	1.531234	1.98502
Neuroblast differentiation-associated protein AHNAK	Q09666	1.527779	2.161916
cDNA, FLJ95010, highly similar to Homo sapiens Bloom syndrome (BLM), mRNA	B2RAN0	1.509638	2.962984
Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	Q9H1E3	1.506829	2.264606
E3 ubiquitin-protein ligase RBBP6	Q7Z6E9	1.501486	2.533098
Interferon regulatory factor 2-binding protein 2	Q7Z5L9	1.499756	3.926243
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	A0A024RAD5	1.499171	2.458352
Host cell factor 1	A6NEM2	1.496914	2.672996
Zinc finger Ran-binding domain-containing protein 2	A0A0C4DGV5	1.476376	2.494644
Mediator of DNA damage checkpoint protein 1	Q14676	1.473804	2.688164
HCG19665, isoform CRA_a	A0A024QZF1	1.461629	2.130096
NF-kappaB repressing factor	A3F768	1.457391	2.227948
Protein PRRC2A	P48634	1.452344	1.740385
AF4/FMR2 family member 4	Q9UHB7	1.417857	3.253733
Leukocyte receptor cluster member 8	A0A087WUE4	1.416961	2.95401
Splicing factor, arginine/serine-rich 15	O95104	1.390012	2.213113
cDNA, FLJ96580, highly similar to Homo sapiens hepatoma-derived growth factor (high-mobility group protein 1-like) (HDGF), mRNA	B2RDE8	1.362928	2.720004
CD2-associated protein	Q9Y5K6	1.357271	2.172356
Lysine-rich nucleolar protein 1	Q1ED39	1.324569	1.586402
cDNA FLJ37346 fis, clone BRAMY2021310, highly similar to Transcriptional repressor p66 beta	B3KSZ4	1.319931	2.803249
U2 snRNP-associated SURP motif-containing protein	O15042	1.316139	2.360484

Msx2-interacting protein	Q96T58	1.312439	2.899443
Ribosomal protein S19 (Fragment)	Q8WVX7	1.309227	2.162313
Cyclin-dependent kinase 12	Q9NYV4	1.306797	1.441003
Myb-binding protein 1A	Q9BQG0	1.305222	2.680913
Regulation of nuclear pre-mRNA domain-containing protein 2	Q5VT52	1.296972	3.497621
cDNA FLJ75211, highly similar to Homo sapiens ubiquitin specific peptidase like 1, mRNA	A8K1B1	1.290854	1.756735
Probable JmjC domain-containing histone demethylation protein 2C	Q15652	1.28818	3.662465
TNPO2 variant protein (Fragment)	Q4LE60	1.283974	2.568543
Double-strand break repair protein MRE11A	P49959	1.283543	3.781323
Origin recognition complex, subunit 2-like (Yeast), isoform CRA_a	A0A024R411	1.27295	2.53586
Splicing factor 3B subunit 2	Q13435	1.272198	2.776691
ESF1 homolog	Q9H501	1.255969	3.863564
Tumor suppressor p53-binding protein 1	Q12888	1.252024	2.029592
Cactin	Q8WUQ7	1.251818	2.613322
Eukaryotic translation initiation factor 5	P55010	1.235951	1.625504
Lysozyme C	P61626	1.22756	1.359255
Serine/threonine kinase 38, isoform CRA_a	A0A024RD18	1.218952	2.672397
cDNA, FLJ93224	B2R713	1.21609	1.304024
Nuclear fragile X mental retardation protein interacting protein 2	A1L3A7	1.2153	1.331827
Shugoshin-like 2	Q562F6	1.192524	1.977134
Poly(U)-binding-splicing factor PUF60	Q9UHX1	1.190017	2.329497
cDNA FLJ53206	B4DUT2	1.188667	2.401761
cDNA, FLJ95650, highly similar to Homo sapiens karyopherin (importin) beta 1 (KPNB1), mRNA	B2RBR9	1.181074	1.901562
FB19 protein	Q2L6I0	1.174618	4.155342
Sex comb on midleg-like protein 2	Q9UQR0	1.169147	2.542609
RNA polymerase-associated protein RTF1 homolog	Q92541	1.158275	3.185639
Coatomer subunit gamma-2	Q9UBF2	1.155413	2.637598

La-related protein 1	Q6PKG0	1.149616	3.053881
Transcription initiation factor TFIID subunit 1	P21675	1.149306	3.599869
Ubiquitin-specific protease 7 isoform (Fragment)	Q6U8A4	1.146947	2.863499
Lysine-specific demethylase 3A	F5H070	1.146464	2.782132
Serine/arginine repetitive matrix 1 isoform 2 (Fragment)	A0A0S2Z4Z6	1.123029	1.772964
Pre-mRNA-processing factor 40 homolog A	P19338	1.111365	2.79678
MKL/myocardin-like protein 2	O75400	1.098342	1.642213
cDNA FLJ16777 fis, clone BRHIP2029567, highly similar to Cell division cycle 5-like protein	Q9ULH7	1.081925	1.449427
Nuclear pore complex protein Nup98-Nup96	B3KY60	1.081469	2.377513
WD repeat-containing protein 43	P52948	1.073947	3.756649
Peptidyl-prolyl cis-trans isomerase-like 4	Q15061	1.068626	2.265753
Zinc finger CCCH domain-containing protein 11A	Q8WUA2	1.062828	2.930983
Transcription elongation regulator 1	O75152	1.061671	1.561305
Translocase of inner mitochondrial membrane 50 homolog	O14776	1.059651	2.340964
CCAAT/enhancer-binding protein zeta	A0A024R0M6	1.05109	2.715821
Microfibrillar-associated protein 1	Q03701	1.043362	2.429041
ATP-dependent RNA helicase DDX42	P55081	1.034498	1.637678
Trinucleotide repeat-containing gene 18 protein	Q86XP3	1.033659	2.35692
Uncharacterized protein C17orf47	H9KVB4	1.031234	1.784598
Death-inducer obliterator 1	Q8NEP4	1.025556	1.88081
cDNA FLJ76656, highly similar to Homo sapiens scaffold attachment factor B (SAFB), mRNA	Q9BTC0	1.019607	2.955056
Probable ATP-dependent RNA helicase DDX10	A8K329	1.012194	2.627393
Importin subunit alpha	Q13206	1.010095	2.010998
Pinin	A0A024RDV7	0.99676	1.577388
Ataxin-2-like protein	Q9H307	0.993746	3.264844

Coiled-coil domain-containing protein 94	Q8WWM7	0.981437	2.436901
Zinc finger protein 638	Q9BW85	0.96992	4.663891
PHD finger protein 6	Q14966	0.96139	2.10817
Microtubule-associated protein 1B, isoform CRA_b	A0A0D9SGE8	0.961353	2.945796
Sentrin-specific protease 6	A0A024RAM4	0.934588	2.872377
cDNA, FLJ94609	Q9GZR1	0.917343	2.605648
28 kDa heat- and acid-stable phosphoprotein	B2R9Y2	0.911135	3.786661
Putative uncharacterized protein DKFZp686I05169	Q13442	0.899499	2.295289
cDNA FLJ78753, highly similar to Homo sapiens zinc fingers and homeoboxes 3 (ZHX3), mRNA	Q6MZP3	0.893272	1.368506
Thyroid receptor-interacting protein 6	A8K8Q0	0.892876	2.026548
Serine arginine-rich pre-mRNA splicing factor SR-A1, isoform CRA_a	Q15654	0.890137	1.910481
Protein capicua homolog	A0A024QZH6	0.881083	3.234565
Histone deacetylase	I3L2J0	0.875791	2.250333
RNA polymerase II-associated factor 1 homolog	Q6IT96	0.872574	1.315722
ZNF521 protein (Fragment)	Q8N7H5	0.864935	2.153751
cDNA FLJ59238, highly similar to SNW domain-containing protein 1	Q8IYZ2	0.862354	1.414091
Pleiotropic regulator 1	B4DEG7	0.861419	3.771498
Protein FAM192A	O43660	0.861415	2.34135
Protein PRRC2C	Q9GZU8	0.855652	1.546051
Replication protein A 70 kDa DNA-binding subunit	Q9Y520	0.832744	1.656849
Insulin receptor substrate 4	P27694	0.799467	2.128269
Squamous cell carcinoma antigen recognized by T-cells 3	O14654	0.797072	1.653278
Protein kinase C-binding protein 1	Q15020	0.791573	2.749668
Ubiquitin-associated protein 2-like	Q9ULU4	0.786746	2.724188
Cytoplasmic FMR1-interacting protein 2	Q14157	0.786108	1.853979
Pachytene checkpoint protein 2 homolog	Q96F07	0.784148	1.855357

Fragile X mental retardation autosomal homolog variant p2K	Q15645	0.782367	2.88105
Midasin	A0A0F7KYT8	0.780846	2.101795
Tubulin alpha-1A chain	Q9NU22	0.759815	2.588918
Serine/arginine repetitive matrix protein 2	Q71U36	0.755911	3.08486
Zinc finger protein 687	Q9UQ35	0.753527	2.968319
Peroxisome proliferator activated receptor interacting complex protein	Q8N1G0	0.752799	1.655331
cDNA, FLJ95388, highly similar to Homo sapiens step II splicing factor SLU7 (SLU7), mRNA	E1NZA1	0.744879	1.951224
DNA polymerase eta	B2RBA0	0.724464	1.823501
Tubulin alpha-1C chain	Q9Y253	0.722088	1.434936
ATP-dependent RNA helicase A	F5H5D3	0.717359	2.492
Nucleophosmin isoform 2	Q08211	0.704302	3.308272
THO complex subunit 2	A0A0S2Z491	0.692535	3.00997
Heterogeneous nuclear ribonucleoprotein L	Q8NI27	0.68892	2.391013
14-3-3 protein theta	P14866	0.685394	1.990324
Histone acetyltransferase type B catalytic subunit	P27348	0.670576	3.368625
TIP41-like protein	O14929	0.666057	1.452336
PRP4 pre-mRNA processing factor 4 homolog B	O75663	0.665711	2.327751
Parafibromin	A0A024QZY5	0.647306	2.93022
Histone-binding protein RBBP7	Q6P1J9	0.646723	2.033434
cDNA FLJ75831, highly similar to Homo sapiens exportin, tRNA (nuclear export receptor for tRNAs) (XPOT), mRNA	Q16576	0.644728	1.65272
Survival motor neuron protein	A8KA19	0.644333	2.395685
General transcription factor 3C polypeptide 5	E7EQZ4	0.639785	2.350308
Biorientation of chromosomes in cell division protein 1-like 1	Q9Y5Q8	0.634204	2.959526
cDNA FLJ56176, highly similar to Poly(A) polymerase alpha	Q8NFC6	0.632562	1.844291

cDNA FLJ33964 fis, clone CTONG2019029, highly similar to Pseudouridylate synthase 7 homolog	B4DYF4	0.627714	2.093342
Beta 5-tubulin	B3KRB2	0.605685	1.904993
Signal transducer and activator of transcription 1-alpha/beta	Q5SU16	0.603546	2.306446
Protein CASC5	P42224	0.595479	1.852176
cDNA, FLJ94025, highly similar to Homo sapiens tripartite motif-containing 28 (TRIM28), mRNA	Q8NG31	0.594615	1.993436
Small acidic protein	B2R8R5	0.587359	1.675933

To determine whether there were similarities between the proteins detected in the ZNS5-FLAG pulldown and those detected in the ZNS5-BirA Streptavidin pulldown, a list of all ZNS5-FLAG interactors from the MaxQuant and PD datasets was constructed within Microsoft Office Excel and compared to the high confidence proteins detected in the streptavidin pulldown (available on the CD provided in Appendix C; filename 'masterlists'). This analysis showed that there were only 8 proteins that were common between the streptavidin pulldowns and the ZNS5-FLAG pulldowns. These are shown in Table 24 and 25 with a comparison between their ratios in the FLAG and streptavidin pulldowns and their function/description according to Uniprot.

Table 24: *Common proteins between all ZNS5 pulldown high confidence proteins*

Protein name	Accession	Function/description	log2 t-test difference (FLAG)	log2 t-test difference (BirA)
Pre-mRNA-processing factor 40 homolog A	O75400	Suppressor of cytoplasmic action of WASL/N-WASP by preventing export from the nucleus	0.693049113	1.098342
RNA-binding protein 25	P49756	Regulator of pre-mRNA splicing and of BCL2L1 expression	0.565931	1.864361
28kDa heat- and acid-stable phosphoprotein	Q13442	Enhances the level of fibroblast growth in response to PDGFA and inhibits mitosis in response to PDGFB	1.656132	0.899499

Table 25: *Common proteins between ZNS5-BirA high confidence interactors and those not found in all three ZNS5-FLAG high confidence interactor datasets*

Protein name	Accession	Function/description	log2 t-test difference (FLAG)	log2 t-test difference (BirA)
Thymopoetin, isoform CRA_c	A0A024RBE7	Plays a structural role within the nucleus and predicted to be involved in post-mitosis nuclear lamina assembly	0.513537963	2.110509
cDNA FLJ77796	A8K8Y8	Has no known role but has mRNA highly similar to evolutionarily conserved G-patch domain containing protein mRNA	1.265382409	1.858805
Tubulin alpha-1C chain	F5H5D3	Major component of microtubules	1.247957865	0.717359
Treacle protein	Q13428	Regulator of RNA Pol I activity	0.517242432	3.046453
Little elongation complex subunit 1	Q9Y2F5	Component in the little elongation complex which is required by RNA Pol II and III to generate snRNA	5.582927386	2.395516

To investigate why there was such a lack of correlation between the two experiments at the individual protein level, the streptavidin pulldown dataset was analysed by STRING and DAVID to determine whether proteins participating in similar pathways were enriched in the pulldown samples. Figure 46 shows the STRING pathway analysis with DAVID functional annotation clusters being shown in Table 26 (also highlighted in Figure 46).

Figure 46:

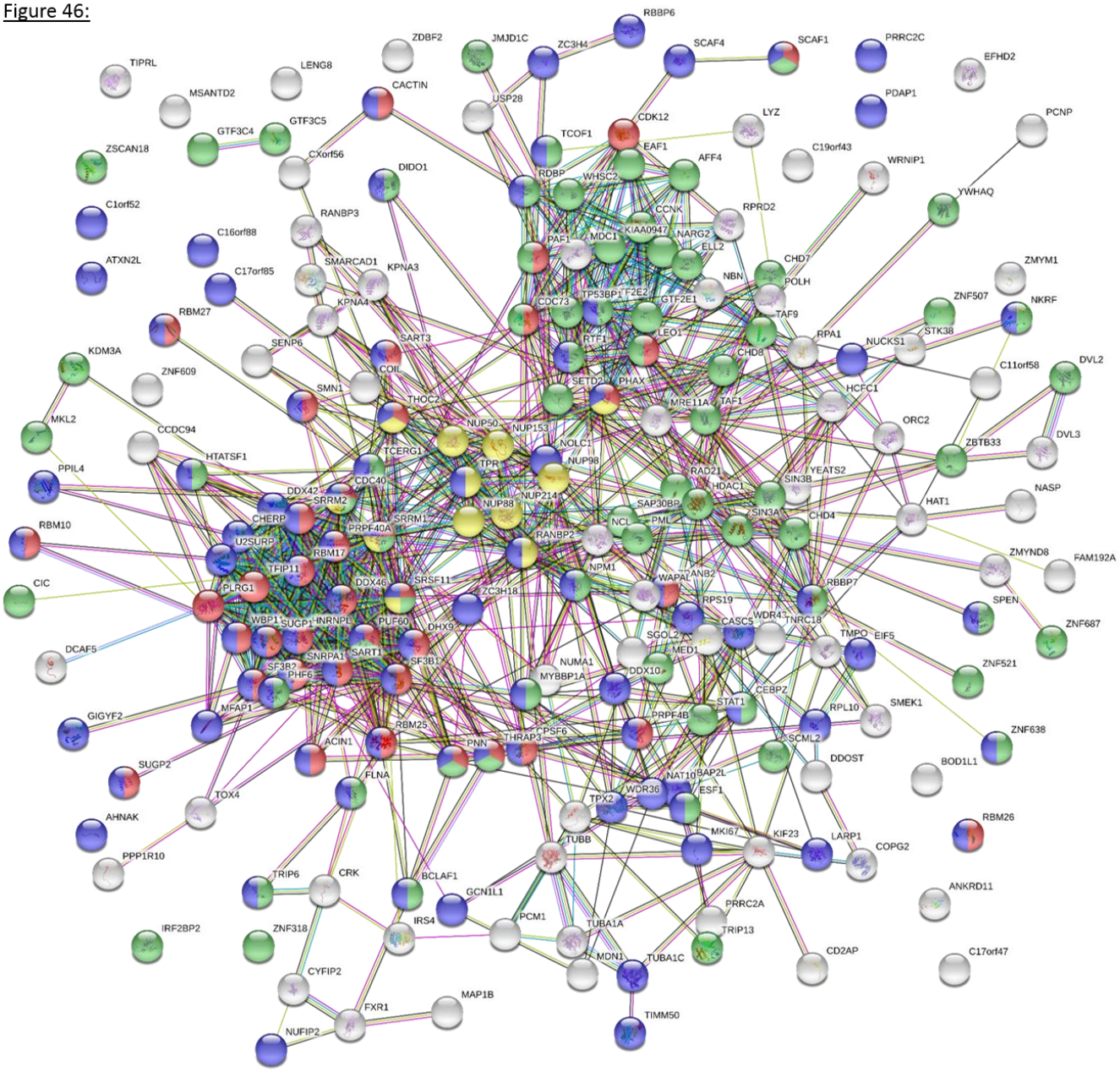

Figure 46: *STRING pathway analysis results on ZNS5-BirA interactors in MaxQuant 0.01FDR* – High confidence interactors of ZNS5 generated from a ZNS5-BirA/streptavidin pulldown presented in STRING with top functional annotation clusters from DAVID analysis highlighted; mRNA processing (red), RNA binding (blue), transcription, DNA templated (green), and RNA transport (yellow)

Table 26: *DAVID/STRING functional annotation clustering for high confidence interactors of ZNS5-BirA*

Functional annotation cluster	DAVID enrichment score	STRING false discovery rate
mRNA processing/splicing/spliceosome	13.26	5.14e-25
RNA-binding	7.68	2.43e-42
Transcription/transcription regulation	6.74	8.41e-14
Cell-cell adhesion	4.48	-
RNA transport/mRNA export from nucleus	4.11	4.79e-06

Cell-cell adhesion was not identified within STRING as it was below the 0.05 false discovery rate cut-off. ‘Ribosome/translation’, the functional cluster which had been the highest enriched in all previous samples was not present at all in the DAVID or STRING analysis of the streptavidin pulldown. An exhaustive list of functional annotation clusters for both the streptavidin pulldown and the FLAG pulldown have been presented in table 27.

Table 27: *Functional annotation clusters for ZNS5-FLAG and ZNS5-BirA high confidence interactors*

ZNS5-FLAG	EASE score	ZNS5-BirA	EASE score
Ribosome/ translation	17.16	mRNA processing/ splicing/ spliceosome	12.07
Histone/ chromatin	13.1	RNA-binding	8.33
Unfolded protein binding/ microtubule	3.95	Transcription/ transcription regulation	7.2
Mitochondrial protein/ inner membrane	3.36	Cell-cell adhesion	4.48
Actin	2.96	RNA transport/mRNA export from nucleus	4.11

Protein biosynthesis/ RNA transport	2.78	G-patch domain	4.10
Nucleotide binding/ ATP binding	2.32	Cell cycle/ cell division	4.07
Unfolded protein binding/ DnaJ	2.03	CID domain	3.21
Spliceosome	1.3	Zinc finger	2.82
Cell cycle/ division	1.3	DNA damage/ DNA repair	2.82
DNA damage/ DNA repair	1.3	Histone modification	2.8
		Centromere	2.68
		Histone deacetylation	2.6
		Nuclear localisation signal binding	2.26
		RAN GTPase binding	1.93
		Helicase	1.87
		Transcription factor TFIID complex	1.82
		Homologous recombination/ ds break repair	1.57
		SAP domain	1.54

The streptavidin pulldown instead showed an enrichment of highly nuclear functional annotation clusters (table 24) and lacked the enrichment of almost all of the individual proteins that were shown in the FLAG pulldown. The lack of correlation between the two pulldowns and the

novelty of the ZNS5-BirA cell lines and streptavidin pulldown protocol influenced the decision on validation targets.

3.7: Validation

3.7.1: Validation targets

The ZNS5 interaction partners chosen for validation were cyclin-dependent kinase 1 (Cdk1) and E3 ubiquitin-protein ligase UBR5 (UBR5). Both were found within the FLAG pulldown datasets with UBR5 being identified within the high confidence interactors list by the MaxQuant 0.05FDR and PD analyses and Cdk1 being detected in the outputs from all three analyses. Both were also ZNS5-specific interactors which afforded the chance to show differences in the interactomes of both ZIKV and DENV. The log2 t-test difference and –log10 p values are presented in table 28.

Table 28: *Cdk1 and UBR5 log2 t-test differences and –log10 p values for ZNS5, D2NS5, and D4NS5 FLAG pulldowns*

Protein	log2 t-test diff. ZNS5	-log10 p value ZNS5	log2 t-test diff. D2NS5	-log10 p value D2NS5	log2 t-test diff. D4NS5	-log10 p value D4NS5
Cdk1	1.128808041	1.241630291	0.329683378	1.473690918	0.104311436	0.200287243
UBR5	1.138352076	1.706456547	-0.10303892	0.817535146	-1.14451909	1.223404439

Cdk1 was a target of interest because control of cell cycle regulation and cell cycle arrest had already been shown to be a characteristic of ZIKV infection (209). Cdk1 is involved with progress through the G2 and M phases of the cell cycle by interacting with cyclin A and B. The targeting of cell cycle regulation is known to occur in other viral infections and whilst in many cases viral infection acts to encourage cell cycle progression, viral infection is also known to prevent cell cycle progression such as the case of SARS-CoV (210). Cdk1 has also been shown to be linked to apoptosis, another key characteristic of ZIKV infection (211). Cdk1 was identified using 10 peptides, 6 of which were unique peptides, giving additional confidence about its identification.

UBR5 was of interest because STAT2 is known to be degraded by ZNS5 in a fashion similar to DENV-NS5 which involves the use of UBR4 to ubiquitin-label STAT2 and target it for proteasomal degradation. UBR4 however was shown previously to not be responsible for this function in ZIKV infection. Proving whether UBR5 interacted with ZNS5 was therefore enticing as the close relation of ZIKV and DENV as well as their similar methods of STAT2 degradation may well have resulted in the use of two E3 ubiquitin-protein ligases from the same family. Within the dataset, it was identified using 10 peptides, 2 of which were unique peptides.

3.7.2: Validation experiments

To validate the differential binding of Cdk1 and UBR5 to ZNS5 and DENV-NS5, reverse pulldowns were performed using antibodies against UBR5 and Cdk1 (details in Table 4 in 2.3.12) with lysates generated from 293-Flp, D2/4NS5-FLAG, and ZNS5-FLAG cells which had been exposed to doxycycline (1µg/ml) for 4d. The pulldowns were then done on the lysates using the specific antibodies and Protein G beads as described in 2.3.11. After the final wash step, proteins were eluted from the beads by heating to 95 ºC for 10 min. The eluate was recovered and kept while the beads were discarded. Western blots were then performed using these samples probing with the anti-FLAG antibody to detect whether ZNS5 or D2/4NS5 were present within the final samples.

Figure 47:

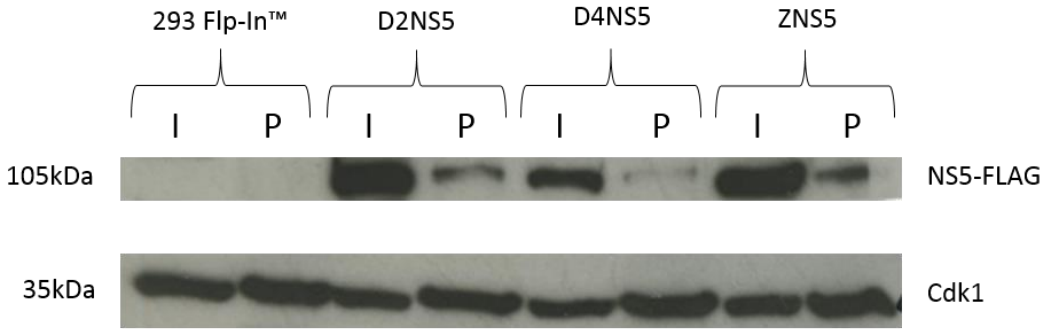

Figure 47: *Western blot showing pre-and post-Cdk1 pulldown samples probed for Cdk1 and FLAG* – Large quantities of 293 Flp-in™, D2NS5, D4NS5, and ZNS5-FLAG cells were grown and lysed. A small sample (*I*) was taken to be compared with the final sample (*P*) after a pulldown was performed with anti-Cdk1 antibody and Protein G agarose beads. Western blot analysis of the samples was performed using anti-FLAG and anti-Cdk1 antibody and developed using a Konica film developer.

As Figure 47 shows, the Cdk1 pulldown was deemed successful with Cdk1 being detected in both pre- and post-pulldown samples. The anti-FLAG antibody detected a band at the size expected for NS5-FLAG, not only for ZNS5 but also to a lesser extent for D2NS5 and an even lesser extent for D4NS5. The result confirmed the proteomic analysis as ZNS5-FLAG had the highest enrichment of Cdk1 with D2NS5 also showing a smaller enrichment although that was not statistically significant in MaxQuant 0.01FDR and PD datasets (table 28).

Due to the size of UBR5 (~309 kDa), wet transfer was used for the Western blot procedure in an attempt to detect UBR5 in the pulldown samples, however it was not possible to obtain a successful western blot for UBR5 before the project ended. Despite this, the samples were probed with the anti-FLAG antibody to investigate whether there was any indication of ZNS5-FLAG present in the post-pulldown samples. Figure 48 shows the anti-FLAG probes of the UBR5 pulldown samples.

Figure 48:

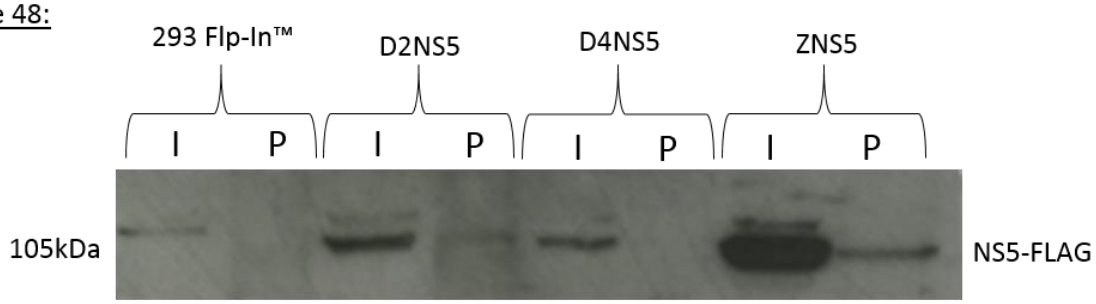

Figure 48: *Western blot showing pre-and post-UBR5 pulldown samples probed for FLAG* – Large quantities of 293 Flp-in™, D2NS5, D4NS5, and ZNS5-FLAG cells were grown, incubated with doxycycline (1µg/ml) for 4 days, and lysed. A small sample was taken (*I*) to be compared with the final sample (*P*) after a pulldown was performed with anti-UBR5 antibody and Protein G agarose beads. Western blot analysis of the samples was performed using anti-FLAG antibody and developed in a Konica film developer.

Figure 48 shows that ZNS5-FLAG was detectable within the UBR5 post-pulldown samples whereas D4NS5-FLAG was not. D2NS5-FLAG was present at a very faint level in the post-pulldown sample. However there was a clear loading difference within, at least, the pre-pulldown samples. The proteomic analysis showed UBR5 was enriched significantly within ZNS5-FLAG FLAG-pulldowns however it showed no significant changes within D2NS5 or D4NS5 samples (table 28).

4: Discussion

4.1: The importance of a direct comparison of NS5 interactors

Currently there is only one study that reports a large scale interactomic analysis of multiple NS5 proteins. Using Y2H analysis, Le Breton *et al* (143), investigated the cellular interaction partners of NS3 and NS5 for multiple flaviviruses, however an extensive comparison between the different flavivirus proteins (for either NS3 or NS5) was not done, instead pooling all flavivirus NS5 interactors from the tested species. Multiple individual NS5 interactomic studies have been performed using; computational prediction (212, 213), Y2H (143, 170, 214), and tagged co-IP (171, 215) however none of these studies been done in tandem with two flavivirus species or serotypes of a single flavivirus. Past studies have been used collectively to discover common interactors for a specific NS5, but each study can have profound experimental differences to another that affect the reliability of being able to compare two individual studies, even when using the same technique. The direct study reported here, using the same methods and side-by-side processing of samples grants a much more robust comparison than previously possible across different studies. It also used the approach of quantitative proteomics rather than using approaches that provide purely qualitative data. Alongside this, the direct comparison of the same samples using different proteomics programmes at different settings and for the same NS5 protein using two different methods of pulldown-proteomics granted insight into the reliability of comparing different studies. The use of stably transformed 293-Flp cells and co-IP LC-MS/MS meant that the NS5 proteins could be expressed at more physiologically relevant levels and naturally localised, unlike in previous Y2H studies.

4.2: ZNS5 damages cell health

Analysis of the stably transfected cell lines produced in this study by western blotting (Figure 17) and IFA analyses (Figures 16 and 18), suggested that the cells expressed ZNS5 as intended. However it was observed that there was a drop in cell viability on day 6 and 8 after doxycycline induction of ZNS5 (as evidenced by the phase contrast images of the cells used in the IFA in Figure 16) that was not seen for the DENV-NS5 cell lines (data not shown). The cells became rounded and grew very slowly, as a population, beyond 4 days of doxycycline exposure, along with a number of them becoming detached, a sign of 293-Flp cell death. At first, it was assumed that perhaps high expression levels of ZNS5 caused cellular toxicity, but the recent literature has repeatedly reported that apoptosis and cell cycle cessation at a sub-G1 phase is a hallmark of ZIKV infection, suggesting that NS5 may play an important role in this process (54). The later discovery that ZNS5 interacted with proteins involved in cell cycle regulation/apoptosis, as well as DNA damage/repair pathways (another hallmark of ZIKV infection due to the formation of pyknotic nuclei) suggested that at least some of the damage to cell health that occurs in ZIKV infection appears to be due to the action of the NS5 protein. A recent study on the interactomics and

transcriptional effects of individual ZIKV proteins, to investigate their apoptotic effect, did not include NS5. This was due to the lack of apoptotic phenotype seen in one of the study's early experiments, however this is in direct opposition to what was seen for the ZNS5-FLAG cell line developed in this study (216). The previous study, however, used fission yeast cells instead of human cells, which are far less biologically relevant and may have lacked key interactors and it also reported that NS5 had no distinct localisation within the yeast cells as seen in IFA analysis which goes against both previous literature (217, 218) and what was seen in Figures 16 and 18.

4.3: The validity of the ZNS5-BirA cell line

The lack of an effective antibody recognising the ZNS5-BirA protein (as shown by Figures 21, 22, and 23) meant that direct detection of ZNS5-BirA in 293-Flp cells to validate ZNS5-BirA protein production was not possible. A functional biotinylation assay (Figure 24) and RT-PCR (Figure 25 and Appendix B) were used to prove that the whole gene was being transcribed and that the ZNS5-BirA protein was produced. As Figure 24 shows, biotin labelling only occurred when ZNS5-BirA cells were incubated with both doxycycline and biotin, which suggested that the BirA tag was being expressed and functioning as expected. However the lack of an effective antibody recognising ZNS5-BirA precluded protein localisation studies. The D2NS5-FLAG and D4NS5-FLAG cell lines have previously been used for NS5 localisation studies (159). As such the ZNS5-FLAG cell line was analysed by IFA (Figure 18) which not only tested the control of the doxycycline-induced expression, but also confirmed that the ZNS5-FLAG protein localised as previously reported for ZNS5 expressed in Vero and HEK-293 cells (217, 218). As epitope tags and modifications can alter protein localisation, it was important to check that the localisation of ZNS5-BirA was unaffected in the new cell lines. Due to the lack of functional antibody, this was not possible to do before the end of the project. Therefore, this was one of the factors taken into account when targets identified by proteomic analysis were selected for validation. As the cellular localisation of ZNS5-FLAG could be confirmed, the interaction partners identified from pulldown using ZNS5-FLAG were given precedence for validation. The ZNS5 protein was recently expressed as a BirA fusion in HEK-293 cells by Coyaud *et al* (218). However the FLAG tag epitope was also fused to the protein, allowing IFA confirmation of ZNS5 protein localisation to the nucleus and cytoplasm. Whilst the BirA tag can clearly be added to the protein without affecting localisation, it was not possible to conclusively state this for the ZNS5-BirA produced in this study.

4.4: Reliability of FLAG and streptavidin pulldowns based on bioinformatic analysis

The proteomic datasets produced using the anti-FLAG antibody and streptavidin pulldown samples showed a high correlation, in terms of the relative amounts of identified proteins between replicate sample groups, as shown by the scatterplot graphs and accompanying Pearson's Correlation values (Figures 32-34, 45) with less correlation seen between the data produced using

samples derived from cells expressing different NS5 proteins (data not shown), as would be expected. The protein search list provided to MaxQuant and PD for protein identification included the Uniprot human proteome and the sequences of the ZNS5, D2NS5, and D4NS5 proteins used in this study. The respective viral NS5 proteins were amongst the top hits in each corresponding dataset, confirming the presence of the proteins in each pull-down sample and their identification by LC-MS/MS and MaxQuant/PD analysis. The identification of the NS5 proteins added confidence to the results, as western blotting (Figure 26) had already shown the success of the pulldowns in enriching the NS5-FLAG proteins in the samples. Proteomic analysis of the streptavidin pulldown samples also showed an enrichment of the ZNS5-BirA protein. As BirA acts as a proximity-labelling tag, it could have biotinylated the ZNS5 protein it was fused to or other molecules of ZNS5-BirA that had localised to the same area intracellularly.

The identification of known NS5 interactors in the datasets is one means to provide validity to the results. STAT2 is the most well characterised interaction partner of the DENV-NS5 and ZNS5 proteins (88, 168) and was previously found to interact with high-confidence with DENV-2 and -4 NS5, using the same cell lines used in this study in a SILAC based proteomic analysis (Dr A. Davidson, personal communication). However, puzzlingly, STAT2 was not identified as an interaction partner for any of the NS5 proteins examined in this analysis. It may have been that STAT2 was targeted for degradation by NS5 within the cells before the pulldown experiments took place. Alternatively, differences in the TMT labelling and SILAC based proteomic approaches may result in different interaction datasets. Another known interactor of DENV-2 NS5 is ELKS/ERC1, which has been identified in numerous NS5 interaction studies (143) including the previous SILAC based study in the laboratory, albeit not as reproducibly as for STAT2. However, once again, this interaction partner was not identified in the pull-down samples for DENV-2 or the other NS5 proteins. It is unclear whether ELKS/ERC1 was not present in sufficient amount in the samples or the MaxQuant/PD bioinformatic analysis failed to identify the protein from the peptide spectral data. Nevertheless, a number of specific interactors of DENV-NS5 were identified within the FLAG pulldown samples by proteomic analysis including the; DnaJ proteins, MCM7, YBOX proteins, DDX proteins (171), PPIases (172), Hsp90 (219, 220), and eukaryotic initiation factors (221) which added confidence and reliability to the datasets generated in this study. The questions raised in 4.3 to do with whether the ZNS5-BirA was functioning/localising properly could also explain why that dataset saw none of the key interactors that we might expect to see like endosomal proteins, ER/ER-stress related proteins, and translation/ribosome proteins (218). Cell cycle proteins and RNA/DNA processing proteins were however enriched within the streptavidin pulldowns but due to the nature of proximity labelling, the 18h biotin incubation, and the unconfirmed localisation discussed in 4.3, the specificity of this labelling is questionable.

In addition to the identification of specific proteins, the identification of groups of proteins belonging to specific cellular pathways by gene ontology enrichment and network analysis is

another way to provide confidence in the interactomic results. Individual proteins can be misidentified in MS/MS proteomic analysis but in a holistic view, it is expected that the majority of proteins would be correctly identified. Analysis of the full set of data together, rather than based on individual interactors would therefore give insight into whether proteins involved in the pathways and cellular processes reported to be affected by NS5 were present in the pulldown-proteomics results. The DAVID (in 3.6.2; Tables 9-18) and STRING analyses (presented in Figures 35-43), revealed the proteins associated with functional annotation terms that were enriched in the pulldown samples. Proteins interacting with D2NS5 were enriched for ribosomal proteins as well as proteins participating in the UPR. Proteins in both of these processes have previously been reported as being enriched in DENV-NS5 interaction studies. Proteins interacting with D4NS5 also showed an enrichment of ribosomal proteins as well as RNA processing/splicing proteins which has also been previously identified as a process targeted by DENV-NS5 (171). DENV-NS5 has also been previously shown to interact with proteins participating in pathways involved in the cytoskeleton and in chromatin/histones (212, 215) which was also the case in this study. ZNS5 has been previously linked to cell cycle arrest, DNA damage, and chromatin/histones (213) and so the enrichment of cell cycle regulation proteins, DNA damage/repair proteins, and histones/chromatin-related proteins as determined by DAVID also added confidence to the results of the proteomic analysis (Table 18).

4.5: Differences and similarities between the ZNS5 and DENV-NS5 interactors

Using the full functional annotation clustering provided by the DAVID analysis, it was possible to already identify similarities and differences between the high confidence interactor lists that were generated from the different NS5 pulldown and subsequent TMT-MS/MS experiments. As shown in Table 18, there is a great deal of similarity in the functional annotation clustering analysis of the high confidence interactors of each NS5 protein. For example; ‘ribosome/translation initiation’ is a common and highly enriched functional annotation cluster defined for all three NS5 proteins. However there are also differences. Proteins interacting with D2NS5 included those associated with the functional annotation cluster of ‘rotamase/isomerase’ that was not observed for D4NS5 and ZNS5. Rotamases/isomerases include PPIases (previously mentioned in 4.4) like cyclophilins and other prolyl isomerases which have previously been reported to interaction with D2NS5 (172). Closer analysis of the proteomic data showed that, although D4NS5 and ZNS5 both had rotamases/isomerases within their high confidence interactors list, D2NS5 had a greater number that were significantly enriched. Another clear difference between the interactomic data was the presence of proteins associated with the functional annotation cluster ‘cell-cell adhesion’ in the DENV-NS5 but not ZNS5 interactome. Although not particularly highly enriched, the D2/4NS5 interactomes did include more high confidence interactors that were classed within this cluster; plasminogen activator inhibitor 1 RNA-binding protein and GTP-binding nuclear protein Ran. Not

much is known about the function of the former besides its interaction with cadherin and a potential role in stabilising mRNA but Ran is a GTP switch which has been shown to be a host interactor with other human viruses (222). It has roles in numerous pathways within the cell, most notably the import and export of RNA and proteins from/to the nucleus. Although identified within the 'cell-cell adhesion' cluster, the interaction of the D2/4NS5 proteins with Ran would much more likely be for the manipulation of this host pathway, rather than of cell-cell adhesion.

Non-POU domain-containing octamer-binding protein (NONO), which was identified as an interactor of D4NS5 and not ZNS5 (Table 21), is an upstream regulator of cGAS-STING; a protein which is already known to interact with DENV (223). It was also shown to interact with DENV-2 RNA during another proteomic study (224). Both D4NS5 and ZNS5 have interaction partners that are involved in DNA damage/DNA repair processes that link to ROS damage which is not unexpected, as DENV and other flaviviruses are known to cause oxidative stress in the cells they infect (225). A number of DNA damage/DNA repair proteins that interacted exclusively with D4NS5 were not specifically DNA repair proteins but were instead linked to the process of cell cycle progression. DNA damage does play a role in suppressing cell cycle progression, however cell cycle regulation interference is a common attribute of numerous viral infections and so the higher enrichment of DNA repair/DNA damage proteins as D4NS5 interaction partners may have resulted from an increased level of cell cycle regulation proteins being detected. The D4NS5 interactors did however include ds break-related proteins; the reason for which is unknown.

Unlike D2NS5 and ZNS5, the D4NS5 interaction partners were not enriched in mitochondrial proteins. Multiple DENV serotypes, including DENV-2, are already known to have an effect on mitochondria (223) and so the enrichment of mitochondrial proteins as D2NS5 and ZNS5 interaction partners, albeit weakly, was not unexpected. Proteins interacting with D4NS5 however, failed to show any signs of enrichment for mitochondrial proteins. The enrichment of ATPases and ribosomal proteins as interaction partners for all three NS5 proteins may have been the result of the weak detergent conditions used for the pulldown assays, which failed to break the interactions between multimer constructs such that an interaction with a single protein within the complex would result in the pulldown of numerous attached proteins.

Beyond differences in functional clustering analysis, there were individual protein differences and similarities of interest due to the evidence in the literature associating them with NS5, although the significance of the interaction, as well as the confirmation of the interactions, needs further study. For example, RNA helicases in the DDX protein family are known to interact with D2NS5 (214) and in this study D2NS5 interacted with DDX39 with high confidence which has not been previously described, whereas D4NS5 interacted with DDX55 which is known to restrict the replication of other viruses (226). DDX17; an RNA helicase that is also well-documented for being a target for viruses such as the reported interaction with influenza virus polymerases (227) was enriched in one replicate of the ZNS5 FLAG-pulldown although it did not make the high confidence

interactor list. Its interaction with DDX5 is particularly important within viral infection (228). Therefore, taking the proteomic data at face value would suggest that all three NS5 proteins potentially interact with RNA helicases and seem to predominantly interact with different members of that family.

In addition, both ZIKV and DENV cause oxidative stress as previously mentioned and both D2NS5 and D4NS5 interacted with AOX1 with high confidence; a mitochondrial protein which functions to reduce oxidative stress by removing ROS. Once again, more investigation would be needed to confirm this interaction and the significance and outcome of the interaction if any as higher levels of AOX1 within the pulldown samples could be from a higher level of AOX1 expression in response to oxidative stress caused by DENV-NS5. ZNS5 however, interacted with a greatly reduced level of AOX1 when compared with the two DENV NS5 proteins, although it did still show a slight but not statistically significant enrichment compared to control cells.

Other examples of differences between the two species include ZNS5's apparent higher affinity for COA5; an important mitochondrial assembly factor for cytochrome c oxidase, and the interaction of ZNS5 with SAMHD1 which has previously been implicated in viral disease and yet was not shown to be enriched within either the D2NS5 or D4NS5 pulldown data.

Therefore, despite the differences in individual proteins identified to interact with the NS5 proteins (as shown by tables 6 and 7), a holistic view of the data reveals a high overall level of similarity in the pathways and host cell processes that both the DENV and ZIKV NS5 proteins interact. Even with the inaccuracies that are known to occur with proteomic analysis, the results fit with what may be expected for two closely related species; the same pathways are targeted but with some differences at the individual protein level. Although the ZNS5 interactome may be less reliable, the clear similarities between it and DENV-NS5's high confidence interactome, as well as the presence of proteins previously identified to be viral targets, allowed a level of confidence in the results.

4.6: Streptavidin pulldown compared to FLAG pulldown

As discussed in 4.3 and 4.4, there are limitations to proteomic analysis and there were some questions regarding the reliability of the BioID labelling that was attempted using the ZNS5-BirA cell lines. As previously mentioned, three independently-raised cell lines were used as triplicate replicates to provide more confidence in the analysis. Western blot analysis (figures 24 and 27) showed that the BioID labelling functioned only under the control of doxycycline and when biotin was available. The streptavidin pulldown technique was not previously developed in the laboratory and had to be adapted from previous examples of streptavidin pulldowns used in studies that focused on fractionated samples, rather than whole cell lysates, however the presence of biotinylated proteins within the pulldown samples (as shown in figure 27) gave reason to believe

that the process had worked. However, for optimal results, a cell line expressing ZNS5-BirA with confirmed correct localisation would need to be used.

The data shown in table 10 along with the comparison in table 24 revealed differences between the datasets generated using the ZNS5-FLAG and ZNS5-BirA pulldowns. Proteins identified to interact with ZNS5-FLAG were deemed to have a more widespread localisation within the cell according to the functional annotation clusters that were shown in Table 27, which is consistent with previous literature. By contrast, the proteins interacting with ZNS5-BirA were almost exclusively classified as belonging to nuclear-based functional annotation clusters. The G-patch domain, CID domain, zinc finger domain, FHA domain, and SAP domain (shown in Table 27) are all domains commonly found within RNA/DNA binding proteins and so are generally found in nuclear-localised proteins. A recent study by Coyaud *et al* (218) utilised BioID labelling to identify proteins interacting with ZNS5 within HEK293 cells and found functional annotation cluster enrichment within translation/RNA processing, splicing machinery, and chromatin binding/regulation which is highly comparable to the level of RNA/DNA-processing proteins shown in tables 26 and 27. The same study reported an interaction with the Cajal body within the nucleus as defined by the labelling of coilin; a high confidence interactor in the streptavidin pulldown as shown in table 23.

The streptavidin pulldown would have been expected to show a large overlap with the interactors identified by the FLAG pulldown, with additional protein identifications signifying non-specific, close proximity proteins and transient interactors. However, as tables 24 and 25 report, there were only eight common proteins between the two interaction datasets. Tables 24 and 25 also show that the common proteins were almost exclusively nuclear in their cellular distribution, with the exception of the 28kDa heat- and acid-stable phosphoprotein. As BioID-labelling covalently links biotin to the targeted protein, the lack of purely cytoplasmic interactors suggested that the ZNS5-BirA protein was not functioning/localising comparably to the ZNS5-FLAG protein. The study from Coyaud *et al* also reported an enrichment of endosomal proteins and ER proteins and, although they reported a greater enrichment of nuclear proteins compared with their co-IP pulldown, endosomal and ER proteins were not seen within the proteomic data generated here. It was therefore not possible to confidently compare transient and stable interactors within the scope of this project using the ZNS5-BirA cell line, due to questions regarding the reliability of the localisation/function of the protein.

4.7: Cdk1 and UBR5

The functional annotation cluster analysis performed using DAVID was, in part, used for validation target selection as well as the comparative studies described in 3.7.1. Both Cdk1 and UBR5 are relatively well studied proteins which meant they had well-defined mechanisms of action

and, as antibodies that recognised them were commercially available, they were deemed suitable targets for validation experiments within the timeframe of this project.

The reverse pulldowns using the anti-Cdk1 and anti-UBR5 antibodies were partially successful. Cdk1 was successfully pulled down as shown by western blot (Figure 47), however the large size of UBR5 required a change in the protein transfer protocol which delayed a successful blot beyond the timespan of this project. The pulldown protocol used for UBR5 was the same as that successfully used for Cdk1, however the procedure ultimately required correct recognition of UBR5 by the anti-UBR5 antibody, which could not be confirmed. The commercially available antibody used had, however, been validated for use in the immunoprecipitation protocol used in this study.

The results of the anti-FLAG western blot analysis on the Cdk1 pulldown showed the presence of NS5-FLAG in all three NS5-FLAG samples, implying an interaction with both DENV-NS5 and ZNS5. In contrast, the proteomic data did not identify Cdk1 as a high-confidence interaction partner for either DENV-NS5 protein. As the two procedures were done with different antibodies (ie anti-FLAG vs anti-Cdk1) the results raises questions about the specificity of the interaction of the Cdk1 antibody for the NS5 proteins. DENV infection, unlike ZIKV, does not show signs of cell-cycle cessation although, as described in table 20, Cdk1 does have numerous roles within the cell including interactions with HSP70 family members, which are known to interact with DENV NS5 (229). However, identification of NS5-FLAG within the Cdk1 pulldown western blots for all viruses (Figure 47) did go against the quantitative results seen in the MS/MS proteomics.

If it is taken that the UBR5 pulldown was successful, then the results shown in figure 48 are promising. However, there is a clear loading issue with the NS5-FLAG quantities, at least in the pre-pulldown samples, and this places the result into contention such that it needs to be repeated. ImageJ analysis suggested that ZNS5 does pulldown more UBR5 than D2NS5 with the ZNS5-FLAG band in Figure 48 having a relative density of 1.152 compared with the D2NS5-FLAG band when adjusting for the variation in pre-pulldown levels of NS5-FLAG. This is a far smaller difference than suggested by the quantitative proteomic analysis within the FLAG pulldowns. The wide variation in quantitative data in the proteomic analysis for UBR5 in the two replicates of D2NS5 (-2.95 and -4.13) means it was not possible to draw a reliable conclusion regarding the level of interaction based on proteomic analysis alone, however the western blot and subsequent ImageJ analysis suggested a weaker (but detectable) interaction than for ZNS5. For D4NS5, UBR5 was only identified in one of the pulldown replicates, so it was not possible to determine the level of interaction from the proteomic analysis, although the result of the western blot shown in Figure 48 shows no evidence for a D4NS5-UBR5 interaction.

4.8: The limitations of proteomic data

4.8.1: The general concept of pulldowns/TMT-MS/MS

There are numerous limitations to TMT-MS/MS proteomics. As mentioned in 3.6.1, MS/MS proteomics is based on prediction. Raw MS/MS data is generated and then proteomic programmes assign the peptide fragments to a specific protein or protein group (if a number of protein isoforms are all matched by the peptide) based upon a match between the peptide spectrum and a theoretical spectrum generated *in silico* using a protein search list. The predictive nature of MS/MS proteomics gives rise to the possibility of both false positives and false negatives. Therefore, even with the validations done on the dataset such as those performed in 3.7.2, all proteins of interest would ideally be experimentally validated as interaction partners by complimentary approaches before they can be confidently considered more than just predicted interactors. With the use of a protein search list from a well annotated organism and rigorous MS/MS analysis, it would be expected that a significant portion of the predicted proteins are correctly assigned and pathways analysis using bioinformatic resources such as STRING or DAVID would be a potential next step in analysing cell processes targeted by viral proteins. However, the analysis can still be skewed by the limitation that the pulldown technique, except for the case of the streptavidin/biotin pulldown, doesn't use strong detergents to disrupt non-specific binding or protein complexes as the detergent will disrupt antibody-protein binding. This means that interactomic datasets produced using techniques which employ mild detergents, such as the anti-FLAG pulldowns, contain protein complexes and non-specific binders that can affect the pathways analysis. D2NS5 is known for interacting with ribosomal complexes in tagged pulldowns (215), leading to enrichment of those proteins in STRING and DAVID, as shown in Figures 35-43 and Tables 9-18. However, DENV-NS5 likely only interacts with one or two of those proteins, with the rest being pulled down because of strong protein-protein interactions within the ribosomal complex. The biotin/streptavidin pulldown approach was adopted in this study as a means to identify proteins that bind directly or indirectly but are in close proximity to the BirA tagged protein. The very high binding affinity of biotin for streptavidin allows for the use of detergents strong enough to dissociate protein complexes and any non-specific binders from the captured proteins. The use of the BirA tag as a proximity labelling tag can still lead to the identification of false positives as the biotinylation reaction does not distinguish between the close proximity generated by protein-protein binding and close proximity caused by naturally protein-dense areas. In addition, the use of an 18h incubation period for protein biotinylation means the BirA tag has the potential to come into close proximity with numerous non-interactors. The streptavidin pulldown samples were therefore expected to contain more specific interactors from the ZNS5-FLAG pulldown list as well as extra proteins that would be either proteins that came into proximity through close localisation but no interaction or those that transiently/weakly interacted with ZNS5 and so would have dissociated from ZNS5 in the weak detergent conditions of the FLAG pulldown.

4.8.2: Programmes

Part of this study looked at the differences between a FLAG pulldown processed through both MaxQuant and Proteome Discoverer as well as looking at the differences between 0.01FDR and 0.05FDR within MaxQuant. All of the datasets used the same protein search list to ensure the only difference would be how each programme assigns the peptide fragments it detects. A direct comparison between the differences between the initial data from all three datasets was shown in Table 5 with a comparison of high-confidence interactors obtained after further processing shown in table 6. The differences using the 0.01FDR and 0.05FDR values were as expected, processing with a 0.05FDR yielded a higher number of identifications compared with a 0.01FDR. However, the differences between MaxQuant and Proteome Discoverer were far greater, with the initial protein identifications made with Proteome Discoverer (shown in table 5) containing considerably more proteins than when the data was processed with MaxQuant at both FDRs. The differences were less extreme once the data was further filtered to produce high confidence protein lists (as shown in table 6) however, when the high confidence proteins were compared it was clear that Proteome Discoverer and MaxQuant were detecting largely different groups of proteins as high confidence interactors. Comparisons between the datasets produced using MaxQuant at the two FDRs showed roughly twice as many common high confidence proteins as the comparison between Proteome Discoverer and MaxQuant at either FDR, with only roughly a third of each high confidence list being common across all three datasets. While this increased the confidence in the common proteins identified in all analyses, the results also highlighted how the use of different proteomic software packages influenced the resultant datasets. As discussed in 4.8.1, the concept of TMT-MS/MS proteomics relies on prediction and there will always be some false positives and false negatives identified, but the large extent of the differences occurring due to differences in the bioinformatic analysis is highlighted in the comparison performed in this study with only a small fraction of interactors being picked up reliably between different programmes.

The comparison of proteomic bioinformatic programmes performed in this study does have the limitation that it was not possible to conclusively determine which programme and FDR performed the best. What can be concluded however, is that a stricter FDR can exclude relevant interactors if they are weakly identified and that there definitely is a large difference in the proteomic results depending on which programme is used to process the raw MS/MS data; a difference that is large enough to have an effect even at the level of full data analysis as evidenced by the STRING (Figures 35-43) and DAVID analysis (Tables 9-18). Given the differences in the results of the proteomic analyses the most reliable identifications are undoubtedly those common to all analyses, a statement that has previously been suggested for robust protein identification (218).

4.9: Future directions

There are a number of potential future directions this project could progress in. The most important future step would be the confirmation of the UBR5 pulldown to determine if there is a specific interaction with ZNS5. Next would be to perform FLAG pulldowns once again on the NS5-FLAG cell lines to then determine *via* western blot whether Cdk1 was present in all pulldowns or whether the FLAG proteomic data showed a much more representative reflection of which NS5 proteins interacted with Cdk1. The specificity of that interaction would be important for determining the accuracy of the quantitative aspects of the proteomic analyses. This project has already revealed a notable difference in protein identification between different proteomic programmes and even different FDRs within one single programme and whether there is also an issue with reliable quantitative information would be important to know. Using the same pulldown sample for both western blots and proteomics side-by-side would allow for a better controlled test of the quality of protein identification from MS/MS proteomics.

Another further step, upon confirming the ZNS5/UBR5 pulldown, would be to analyse whether UBR5, ZNS5, and STAT2 all co-localise using IFA imaging as well as performing a UBR5 knock-down in ZNS5 cells to see whether STAT2 levels were affected. In this way it would be possible to determine whether ZNS5 utilises UBR5 to degrade STAT2.

The ZNS5-BirA cell lines showed promise in that the proximity-labelling proved to be successful however without a functioning antibody that can be used to test the localisation of the ZNS5-BirA protein, further work on that cell line raises questions about the validity of the interactions detected. Potentially, the cell line could be remade with the inclusion of a FLAG tag in the ZNS5-BirA fusion, or a more specific antibody could be generated.

Further validations could also be made using reverse pulldowns (as in 3.7.2), siRNA knock-downs, or fluorescent imaging co-localisation analysis to confirm the interaction of other representative proteins within the MS/MS proteomic datasets.

The use of robust controls within experiments is vital for the validity of the findings. However, within this project, numerous experiments did not have the appropriate controls and so repeating those experiments with the proper positive and negative controls would be required. These are as follows:

- Figure 16 and 17 – The introduction of a ‘day 0’ sample showing the expression of ZNS5-FLAG in the absence of doxycycline as well as including a 293-Flp cell line which contains no ZNS5-FLAG gene and a known expresser of the FLAG; either a 293-Flp cell line with solely the FLAG tag introduced into the Flp site or the previously-documented DENV-NS5-FLAG cell lines which have already been shown to express selectively under doxycycline control to ensure the specificity of the FLAG antibody.

Both of these controls would be treated to the same conditions as the samples in terms of doxycycline exposure.

- Figure 21 and 22 – The inclusion of a positive control would greatly strengthen these results but the lack of proven antibody to both ZNS5 and BirA meant that it was not possible. Recreating the cell line using a ZNS5-BirA-FLAG construction would allow for the use of the already-tested FLAG antibody to function as our positive control for both the BirA and ZNS5 antibodies to determine their functionality against.
- Figure 23 – The ZNS5-FLAG '4 days' and '6 days' samples acted as positive controls for this experiment however the failure to include a negative control in the form of lysate from 293-Flp cells treated to 4 days doxycycline as per the test samples C3, C4, and C5 needs to be rectified.
- Pulldowns – In all performed pulldown experiments in this study, the 293-Flp cell line was used as a negative control however it was not the most appropriate nor effective control. A far better control would have been to run simultaneous pulldowns from lysed NS5-FLAG cell lines using antibodies against proteins known to not interact with NS5, like GAPDH, to determine whether the presence of the NS5-FLAG proteins in the pulldowns in Figures 47 and 48 was due to NS5/Cdk1 and NS5/UBR5 interactions as had been originally assumed or non-specific interactions between NS5-FLAG and the antibody-bead constructs. Ideally the antibodies used would be similar to those using in the pulldowns; IgG mouse antibodies for the Cdk1 pulldown and IgG goat for the UBR5 pulldown, to create as close to the same environment as the pulldowns as possible except for the target of the antibody and so the bound protein for NS5 to interact with. In that way, the specificity of the protein-protein binding can be confirmed.

These changes would significantly improve the validity of the findings of this study and therefore allow for more conclusive arguments to be made about the results.

5: References

1. Roehrig JT, Barrett AD. Flavivirus Infections in Humans. eLS. 2013.
2. Organisation WH. WHO | Prevention of sexual transmission of Zika virus. In: WHO, editor. WHO. World Health Organisation Website: World Health Organization; 2016.
3. Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, et al. ICTV Virus Taxonomy Profile: Flaviviridae. J Gen Virol. 2017;98(1):2-3.
4. Schweitzer BK, Center UoNM, Chapman NM, Microbiology P, Iwen PC, piwen@unmc.edu, et al. Overview of the Flaviviridae with an emphasis on the Japanese encephalitis group viruses. Laboratory Medicine. 2017;40(8):493-9.
5. Gubler DJ. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21st Century. Trop Med Health. 2011;39(4 Suppl):3-11.
6. Guzman MG, Kouri GP, Bravo J, Soler M, Vazquez S, Morier L. Dengue hemorrhagic fever in Cuba, 1981: a retrospective seroepidemiologic study. Am J Trop Med Hyg. 1990;42(2):179-84.
7. Gubler D. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. CellPress. 2002;10(2):100-3.
8. Organisation WH. Dengue: Guidelines for diagnosis, treatment, prevention and control. K.Ciceri, P.Tissot, editors: WHO publications; 2009. 147 p.
9. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504-7.
10. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis. 2012;6(8):e1760.
11. Shepard DS, Undurraga ES, Halasa YA, Stanaway JD. The global economic burden of dengue: a systematic analysis - The Lancet Infectious Diseases. Lancet Infectious Diseases. 2017;16(8):935-41.
12. Hadinergoro SR, Arredondo-Garcia JL, Capeding MR, Deseda C, Chotpitayasunondh T, Dietze R, et al. Efficacy and Long-Term Safety of a Dengue Vaccine in Regions of Endemic Disease — NEJM. N Engl J Med. 2015;373:1195-206.
13. Villar L, Dayan G, Arredondo-Garcia JL, Rivera DM, Cunha R, Deseda C, et al. Efficacy of a Tetravalent Dengue Vaccine in Children in Latin America. N Engl J Med. 2015;372:113-23.
14. Aguiar M, Stollenwerk N, Halstead SB. The risks behind Dengvaxia recommendation. The Lancet Infectious Diseases. 2016;16(8):882-3.
15. Ferguson NM. Dengvaxia increases risk for dengue virus in low-transmission areas. Science. 2016.
16. Meltzer M, Rigau-Perez J, Clark G, Reiter P, Gubler D. Using disability-adjusted life years to assess the economic impact of dengue in Puerto Rico: 1984–1994. The American Journal of Tropical Medicine and Hygiene. 1998.
17. Maciel-de-Freitas R, Aguiar R, Bruno RV, Guimaraes MC, Lourenco-de-Oliveira R, Sorgine MH, et al. Why do we need alternative tools to control mosquito-borne diseases in Latin America? Mem Inst Oswaldo Cruz. 2012;107(6):828-9.
18. Dick GWA, Haddow AJ. Zika Virus (I). Isolations and serological specificity. Transactions of The Royal Society of Tropical Medicine and Hygiene. 1952;46(5):509-20.
19. Bearcroft WGC. Zika virus infection experimentally induced in a human volunteer. Transactions of The Royal Society of Tropical Medicine and Hygiene. 1956;50(5):442-8.
20. Marchette NJ, Garcia R, Rudnick A. Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. Am J Trop Med Hyg. 1969;18(3):411-5.
21. Olson JG, Ksiazek TG, Suhandiman, Triwibowo. Zika virus, a cause of fever in Central Java, Indonesia. Trans R Soc Trop Med Hyg. 1981;75(3):389-93.
22. Fagbami AH. Zika virus infections in Nigeria: virological and seroepidemiological investigations in Oyo State. J Hyg (Lond). 1979;83(2):213-9.
23. Ramzy A. Experts study Zika's path from first outbreak in Pacific. New York Times: New York Times; 2016.
24. Xiao-xia G, Chun-xiao L, Yong-qiang D, Dan X, Qin-mei L, Qun W, et al. Culex pipiens quinquefasciatus: a potential vector to transmit Zika virus. Emerging Microbes & Infections. 2016;5(9).

25. Wong SS, Poon RW, Wong SC. Zika virus infection-the next wave after dengue? J Formos Med Assoc. 2016;115(4):226-42.
26. Hart CE, Roundy CM, Azar SR, Huang JH, Yun R, Reynolds E, et al. Zika Virus Vector Competency of Mosquitoes, Gulf Coast, United States. Emerg Infect Dis. 2017;23(3):559-60.
27. Heitmann A, Jansen S, Luhken R, Leggewie M, Badusche M, Pluskota B, et al. Experimental transmission of zika virus by mosquitoes from central Europe. Eurosurveillance. 2017;22(2).
28. Huang YJ, Ayers VB, Lyons AC, Unlu I, Alto BW, Cohnstaedt LW, et al. Culex Species Mosquitoes and Zika Virus. Vector Borne Zoonotic Dis. 2016;16(10):673-6.
29. Arsuaga M, Bujalance SG, Diaz-Menendez M, Vazquez A, Arribas JR. Probable sexual transmission of Zika virus from a vasectomised man. Lancet Infect Dis. 2016;16(10):1107.
30. Osuna CE, Lim S-Y, Deleage C, Griffin BD, Stein D, Schroeder LT, et al. Zika viral dynamics and shedding in rhesus and cynomolgus macaques. Nature Medicine. 2016;22:1448-55.
31. Musso D, Nhan T, Robin E, Roche C, Bierlaire D, Zisou K, et al. Potential for zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, Novemeber 2013 to February 2014. Euro Surveill. 2014;19(14).
32. Goldschmidt D. Sperm donated in Florida may contain Zika, CDC says2017. Available from: http://www.cnn.com/2017/03/13/health/zika-florida-sperm-warning/index.html.
33. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global environmental suitability for Zika virus. Elife. 2016;5.
34. Lee BY, Alfaro-Murillo JA, Parpia AS, Asti L, Wedlock PT, Hotez PJ, et al. The potential economic burden of Zika in the continental United States. PLoS Negl Trop Dis. 2017;11(4):e0005531.
35. Toy S. Brazil risks new Zika outbreak 3 months after end of health emergency. USA Today. 2017.
36. Johansson MA, Mier-y-Teran-Romero L, Reefhuis J, Gilboa SM, Hills SL. Zika and the Risk of Microcephaly. N Engl J Med. 2016;375(1):1-4.
37. Frontera JA, da Silva IR. Zika Getting on Your Nerves? The Association with the Guillain-Barre Syndrome. N Engl J Med. 2016;375(16):1581-2.
38. Siu R, Bukhari W, Todd A, Gunn W, Huang QS, Timmings P. Acute Zika infection with concurrent onset of Guillain-Barre Syndrome. Neurology. 2016;87(15):1623-4.
39. Niemeyer B, Niemeyer R, Borges R, Marchiori E. Acute Disseminated Encephalomyelitis Following Zika Virus Infection. Eur Neurol. 2017;77(1-2):45-6.
40. Zare Mehrjardi M, Keshavarz E, Poretti A, Hazin AN. Neuroimaging findings of Zika virus infection: a review article. Jpn J Radiol. 2016;34(12):765-70.
41. Colombia Zika outbreak: microcephaly cases four times higher this year. The Guardian. 2016 2016-12-09.
42. Chia J. 'Zika is scarier than we initially thought': CDC says. The Daily Mail. 2016 2016-04-11.
43. Fox M. As Zika Spread, Paralyzing Guillain-Barré Syndrome Skyrocketed - NBC News. NBC News. 2016 2016-09-01.
44. Public-Health-England. Zika virus: travel advice. In: England PH, editor. www.gov.uk/guidance2017.
45. Macciocchi D, Lanini S, Vairo F, Zumla A, Figueiredo LT, Lauria FN, et al. Short-term economic impact of the Zika virus outbreak. New Microbiol. 2016;39(4):287-9.
46. Duong V, Lambrechts L, Paul RE, Ly S, Lay RS, Long KC, et al. Asymptomatic humans transmit dengue virus to mosquitoes. PNAS. 2015;112(47):14688-93.
47. Hung NT. Fluid management for dengue in children. Paediatr Int Child Health. 2012;32(s1):39-42.
48. Morell P, Quarles RH. The Myelin Sheath. 6th ed. Siegal G, Agranoff B, Albers R, editors. Philadelphia: Lippincott-Raven; 1999 1999.
49. Hisahara S, Okano H, Miura M. Caspase-mediated oligodendrocyte cell death in the pathogenesis of autoimmune demyelination. Neurosci Res. 2003;46(4):387-97.
50. Alehan FK, Kahveci S, Uslu Y, Yildirim T, Yilmaz B. Acute disseminated encephalomyelitis associated with hepatitis A virus infection. Ann Trop Paediatr. 2004;24(2):141-4.
51. Kanzaki A, Yabuki S. Acute disseminated encephalomyelitis (ADEM) associated with cytomegalovirus infection--a case report. Rinsho Shinkeigaku. 1994;34(5):511-3.
52. Mohsen H, Abu Zeinah GF, Elsotouhy AH, Mohamed K. Acute disseminated encephalomyelitis following infectious mononucleosis in a toddler. BMJ Case Rep. 2013;2013.

53. Sim JE, Lee JB, Cho YN, Suh SH, Kim JK, Lee KY. A Case of Acute Disseminated Encephalomyelitis Associated with Hepatitis C Virus Infection. Yonsei Med J. 2012;53(4):856-8.
54. Garcez PP, Nascimento JM, Vasconcelos JMd, Costa RMd, Delvecchio R, Trindade P, et al. Zika virus disrupts molecular fingerprinting of human neurospheres. Scientific Reports, Published online: 23 January 2017; | doi:101038/srep40780. 2016.
55. Researchers Sound Alarm Over Zika's Potentially Harmful Heart Effects [press release]. Online pubication: American College of Cardiology2017.
56. Sirohi D, Chen Z, Sun L, Klose T, Pierson TC, Rossmann MG, et al. The 3.8 Å resolution cryo-EM structure of Zika virus. 2016.
57. Zhang X, Ge P, Yu X, Brannan JM, Bi G, Zhang Q, et al. Cryo-EM structure of the mature dengue virus at 3.5-A resolution. Nat Struct Mol Biol. 2013;20(1):105-10.
58. Gebhard LG, Filomatori CV, Gamarnik AV. Functional RNA Elements in the Dengue Virus Genome. Viruses. 2011;3(9):1739-56.
59. Rothman AL. Immunity to dengue virus: a tale of original antigenic sin and tropical cytokine storms. Nature Reviews Immunology. 2011;11(8):532-43.
60. Vorou R. Zika virus, vectors, reservoirs, amplifying hosts, and their potential to spread worldwide: what we know and what we should investigate urgently. Int J Infect Dis. 2016;48:85-90.
61. Rodriguez-Roche R, Gould EA. Understanding the Dengue Viruses and Progress towards Their Control. Biomed Res Int. 2013;2013.
62. Paz-Bailey G, Rosenberg ES, Doyle K, Munoz-Jordan J, Santiago GA, Klein L, et al. Persistence of Zika Virus in Body Fluids — Preliminary Report. NEJM. 2017.
63. Russell K, Epidemic Intelligence Service, Influenza Division NCfIaRD, Hills SL, Division of Vector-Borne Diseases a, Oster AM, et al. Male-to-Female Sexual Transmission of Zika Virus—United States, January–April 2016. Clinical Infectious Diseases. 2017;64(2):211-3.
64. Deckard D, Chung W, Brooks J, Smith J, Woldai S, Hennessey M, et al. Male-to-Male Sexual Transmission of Zika Virus. Morbidity and Mortality Weekly Report. 2016(65):372-4.
65. Davidson A, Slavinski S, Komoto K, Rakeman J, Weiss D. Suspected Female-to-Male Sexual Transmission of Zika Virus — New York City, 2016 | MMWR. Mortal Wkly Rep. 2016(65):716-7.
66. Towers S, Brauer F, Castillo-Chavez C, Falconar AK, Mubayi A, Romero-Vivas CM. Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. Epidemics. 2016;17:50-5.
67. Wu SJ, Grouard-Vogel G, Sun W, Mascola JR, Brachtel E, Putvatana R, et al. Human skin Langerhans cells are targets of dengue virus infection. Nat Med. 2000;6(7):816-20.
68. Noisakran S, Onlamoon N, Songprakhon P, Hsiao HM, Chokephaibulkit K, Perng GC. Cells in Dengue Virus Infection In Vivo. Advances in Virology. 2010;2010:15.
69. van der Schaar H, Rust M, Chen C, van der Ende-Metselaar H, Wilschut J, Zhuang X, et al. Dissecting the Cell Entry Pathway of Dengue Virus by Single-Particle Tracking in Living Cells. PLOS Pathogens. 2008.
70. Modis Y, Ogata S, Clements D, Harrison SC. Structure of the dengue virus envelope protein after membrane fusion. Nature. 2004;427(6972):313-9.
71. Holden KL, Harris E. Enhancement of dengue virus translation: role of the 3' untranslated region and the terminal 3' stem-loop domain. Virology. 2004;329(1):119-33.
72. Junjhon J, Pennington JG, Edwards TJ, Perera R, Lanman J, Kuhn RJ. Ultrastructural Characterization and Three-Dimensional Architecture of Replication Sites in Dengue Virus-Infected Mosquito Cells. Journal of Virology. 2014;88(9):4676-97.
73. Fernandez-Garcia MD, Mazzon M, Jacobs M, Amara A. Pathogenesis of flavivirus infections: using and abusing the host cell. Cell Host Microbe. 2009;5(4):318-28.
74. Liu ZY, Li XF, Jiang T, Deng YQ, Ye Q, Zhao H, et al. Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. Elife. 2016;5.
75. Marianneau P, Steffan AM, Royer C, Drouet MT, Jaeck D, Kirn A, et al. Infection of primary cultures of human Kupffer cells by Dengue virus: no viral progeny synthesis, but cytokine production is evident. J Virol. 1999;73(6):5201-6.

76. Thepparit C, Smith DR. Serotype-Specific Entry of Dengue Virus into Liver Cells: Identification of the 37-Kilodalton/67-Kilodalton High-Affinity Laminin Receptor as a Dengue Virus Serotype 1 Receptor. J Virol. 2004;78(22):12647-56.
77. Samanta J, Sharma V. Dengue and its effects on liver. World J Clin Cases. 2015;3(2):125-31.
78. Carr JM, Hocking H, Bunting K, Wright PJ, Davidson A, Gamble J, et al. Supernatants from dengue virus type-2 infected macrophages induce permeability changes in endothelial cell monolayers. J Med Virol. 2003;69(4):521-8.
79. Dung NTP, Le Duyen HT, Van Thuy NT, Van Ngoc T, Van Vinh Chau N, Hien TT, et al. Timing of CD8+ T Cell Responses in Relation to Commencement of Capillary Leakage in Children with Dengue. J Immunol. 2010;184(12):7281-7.
80. Modhiran N, Watterson D, Muller DA, Panetta AK, Sester DP, Liu L, et al. Dengue virus NS1 protein activates cells via Toll-like receptor 4 and disrupts endothelial cell monolayer integrity. Sci Transl Med. 2015;7(304):304ra142.
81. Wu-Hsieh BA, Yen YT, Chen HC. Dengue hemorrhage in a mouse model. Ann N Y Acad Sci. 2009;1171 Suppl 1:E42-7.
82. Pagani I, Ghezzi S, Ulisse A, Rubio A, Turrini F, Garavaglia E, et al. Human Endometrial Stromal Cells Are Highly Permissive To Productive Infection by Zika Virus. Sci Rep. 2017;7:44286.
83. Costa HE, Gouilly J, Mansuy J-M, Chen Q, Levy C, Cartron G, et al. Zika virus reveals broad tissue and cell tropism during the first trimester of pregnancy. Scientific Reports, Published online: 19 October 2016; | doi:101038/srep35296. 2016.
84. Bowen JR, Quicke KM, Maddur MS, O'Neal JT, McDonald CE, Fedorova NB, et al. Zika Virus Antagonizes Type I Interferon Responses during Infection of Human Dendritic Cells. PLoS Pathog. 2017;13(2):e1006164.
85. Quicke KM, Bowen JR, Johnson EL, McDonald CE, Ma H, O'Neal JT, et al. Zika Virus Infects Human Placental Macrophages. Cell Host Microbe. 2016;20(1):83-90.
86. Li F, Wang PR, Qu LB, Yi CH, Zhang FC, Tang XP, et al. AXL is not essential for Zika virus infection in the mouse brain. Emerg Microbes Infect. 2017;6(3).
87. Hamel R, Dejarnac O, Wichit S, Ekchariyawat P, Neyret A, Luplertlop N, et al. Biology of Zika Virus Infection in Human Skin Cells. J Virol. 2015;89(17):8880-96.
88. Grant A, Ponia SS, Tripathi S, Balasubramaniam V, Miorin L, Sourisseau M, et al. Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. Cell Host Microbe. 2016;19(6):882-90.
89. Kumar A, Hou S, Airo AM, Limonta D, Mancinelli V, Branton W, et al. Zika virus inhibits type-I interferon production and downstream signaling. EMBO Rep. 2016;17(12):1766-75.
90. Dowall SD, Graham VA, Rayner E, Atkinson B, Hall G, Watson RJ, et al. A Susceptible Mouse Model for Zika Virus Infection. PLoS Negl Trop Dis. 2016;10(5):e0004658.
91. Lazear HM, Govero J, Smith AM, Platt DJ, Fernandez E, Miner JJ, et al. A Mouse Model of Zika Virus Pathogenesis. Cell Host Microbe. 2016;19(5):720-30.
92. Ma W, Li S, Ma S, Jia L, Zhang F, Zhang Y, et al. Zika Virus Causes Testis Damage and Leads to Male Infertility in Mice. Cell. 2016;167(6):1511-24.e10.
93. Nicastri E, Castilletti C, Liuzzi G, Iannetta M, Capobianchi MR, Ippolito G. Persistent detection of Zika virus RNA in semen for six months after symptom onset in a traveller returning from Haiti to Italy, February 2016. Euro Surveill. 2016;21(32).
94. Bhatnagar J, Rabeneck DB, Martines RB, Reagan-Steiner S, Ermias Y, Estetter LB, et al. Zika Virus RNA Replication and Persistence in Brain and Placental Tissue. Emerg Infect Dis. 2017;23(3):405-14.
95. Singh PK, Guest JM, Kanwar M, Boss J, Gao N, Juzych MS, et al. Zika virus infects cells lining the blood-retinal barrier and causes chorioretinal atrophy in mouse eyes. JCI Insight. 2017;2(4):e92340.
96. Cragan JD MC, Petersen EE, et al. Baseline Prevalence of Birth Defects Associated with Congenital Zika Virus Infection — Massachusetts, North Carolina, and Atlanta, Georgia, 2013–2014 | MMWR. Morbidity and Mortality Weekly Report. 2017;66:219-22.
97. da Silva A, Ganz J, Sousa P, Doriqui M, Ribeiro M, Branco M, et al. Early growth and neurologic outcomes of infants with probably congenital Zika virus syndrome. Emerging Infectious Diseases. 2016;22(11).

98. Ramalho FS, Yamamoto AY, da Silva LL, Figueiredo LT, Rocha LB, Neder L, et al. Congenital Zika virus infection induces severe spinal cord injury. Clin Infect Dis. 2017.
99. Gan ES, Cheong WF, Chan KR, Ong EZ, Chai X, Tan HC, et al. Hypoxia enhances antibody-dependent dengue virus infection. Embo j. 2017;36(10):1348-63.
100. Pardi N, Hogan MJ, Pelc RS, Muramatsu H, Andersen H, DeMaso CR, et al. Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. Nature. 2017;543(7644):248-51.
101. Aliota MT, Dudley DM, Newman CM, Mohr EL, Gellerup DD, Breitbach ME, et al. Heterologous Protection against Asian Zika Virus Challenge in Rhesus Macaques. PLoS Negl Trop Dis. 2016;10(12):e0005168.
102. Iovine NM, Lednicky J, Cherabuddi K, Crooke H, White SK, Loeb JC, et al. Coinfection With Zika and Dengue-2 Viruses in a Traveler Returning From Haiti, 2016: Clinical Presentation and Genetic Analysis. Clin Infect Dis. 2017;64(1):72-5.
103. Castanha P, Nascimento E, Cynthia B, Cordeiro M, de Carvalho O, de Mendonca L, et al. Dengue virus (DENV)-specific antibodies enhance Brazilian Zika virus (ZIKV) infection. Journal of Infectious Diseases. 2016.
104. Boigard H, Alimova A, Martin GR, Katz A, Gottlieb P, Galarza JM. Zika virus-like particle (VLP) based vaccine. PLoS Negl Trop Dis. 2017;11(5):e0005608.
105. Samsa MM, Mondotte JA, Caramelo JJ, Gamarnik AV. Uncoupling cis-Acting RNA elements from coding sequences revealed a requirement of the N-terminal region of dengue virus capsid protein in virus particle formation. J Virol. 2012;86(2):1046-58.
106. Kofler RM, Heinz FX, Mandl CW. Capsid protein C of tick-borne encephalitis virus tolerates large internal deletions and is a favorable target for attenuation of virulence. J Virol. 2002;76(7):3534-43.
107. Schlick P, Kofler RM, Schittl B, Taucher C, Nagy E, Meinke A, et al. Characterization of West Nile virus live vaccine candidates attenuated by capsid deletion mutations. Vaccine. 2010;28(36):5903-9.
108. Samsa MM, Mondotte JA, Iglesias NG, Assuncao-Miranda I, Barbosa-Lima G, Da Poian AT, et al. Dengue virus capsid protein usurps lipid droplets for viral particle formation. PLoS Pathog. 2009;5(10):e1000632.
109. Zhang Y, Corver J, Chipman PR, Zhang W, Pletnev SV, Sedlak D, et al. Structures of immature flavivirus particles. EMBO J. 2003;22(11):2604-13.
110. Rodenhuis-Zybert IA, van der Schaar HM, da Silva Voorham JM, van der Ende-Metselaar H, Lei HY, Wilschut J, et al. Immature Dengue Virus: A Veiled Pathogen? PLoS Pathog. 2010;6(1).
111. Richter MKS, da Silva Voorham JM, Torres Pedraza S, Hoornweg TE, van de Pol DPI, Rodenhuis-Zybert IA, et al. Immature Dengue Virus Is Infectious in Human Immature Dendritic Cells via Interaction with the Receptor Molecule DC-SIGN. PLoS One. 2014;9(6).
112. Mukhopadhyay S, Kuhn RJ, Rossmann MG. A structural perspective of the flavivirus life cycle. Nat Rev Microbiol. 2005;3(1):13-22.
113. Zhang Y, Zhang W, Ogata S, Clements D, Strauss JH, Baker TS, et al. Conformational Changes of the Flavivirus E Glycoprotein. Structure. 2004;12(9):1607-18.
114. Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, et al. Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. Nat Struct Biol. 2003;10(11):907-12.
115. Lee E, Lobigs M. E Protein Domain III Determinants of Yellow Fever Virus 17D Vaccine Strain Enhance Binding to Glycosaminoglycans, Impede Virus Spread, and Attenuate Virulence▽. J Virol. 2008;82(12):6024-33.
116. Op De Beeck A, Molenkamp R, Caron M, Ben Younes A, Bredenbeek P, Dubuisson J. Role of the transmembrane domains of prM and E proteins in the formation of yellow fever virus envelope. J Virol. 2003;77(2):813-20.
117. Keasey SL, Pugh CL, Jensen SM, Smith JL, Hontz RD, Durbin AP, et al. Antibody Responses to Zika Virus Infections in Environments of Flavivirus Endemicity. Clin Vaccine Immunol. 2017;24(4).
118. Chavez JH, Silva JR, Amarilla AA, Moraes Figueiredo LT. Domain III peptides from flavivirus envelope protein are useful antigens for serologic diagnosis and targets for immunization. Biologicals. 2010;38(6):613-8.

119. Chu JJ, Rajamanonmani R, Li J, Bhuvanakantham R, Lescar J, Ng ML. Inhibition of West Nile virus entry by using a recombinant domain III from the envelope glycoprotein. J Gen Virol. 2005;86(Pt 2):405-12.
120. Chu JH, Chiang CC, Ng ML. Immunization of flavivirus West Nile recombinant envelope domain III protein induced specific immune response and protection against West Nile virus infection. J Immunol. 2007;178(5):2699-705.
121. Gutsche I, Coulibaly F, Voss JE, Salmon J, d'Alayer J, Ermonval M, et al. Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. Proc Natl Acad Sci U S A. 2011;108(19):8003-8.
122. Westaway EG, Goodman MR. Variation in distribution of the three flavivirus-specified glycoproteins detected by immunofluorescence in infected Vero cells. Arch Virol. 1987;94(3-4):215-28.
123. Winkler G, Maxwell SE, Ruemmler C, Stollar V. Newly synthesized dengue-2 virus nonstructural protein NS1 is a soluble protein but becomes partially hydrophobic and membrane-associated after dimerization. Virology. 1989;171(1):302-5.
124. Jacobs MG, Robinson PJ, Bletchly C, Mackenzie JM, Young PR. Dengue virus nonstructural protein 1 is expressed in a glycosyl-phosphatidylinositol-linked form that is capable of signal transduction. Faseb j. 2000;14(11):1603-10.
125. Avirutnan P, Fuchs A, Hauhart RE, Somnuke P, Youn S, Diamond MS, et al. Antagonism of the complement component C4 by flavivirus nonstructural protein NS1. J Exp Med. 2010;207(4):793-806.
126. Chung KM, Liszewski MK, Nybakken G, Davis AE, Townsend RR, Fremont DH, et al. West Nile virus nonstructural protein NS1 inhibits complement activation by binding the regulatory protein factor H. Proc Natl Acad Sci U S A. 2006;103(50):19111-6.
127. Lin SW, Chuang YC, Lin YS, Lei HY, Liu HS, Yeh TM. Dengue virus nonstructural protein NS1 binds to prothrombin/thrombin and inhibits prothrombin activation. J Infect. 2012;64(3):325-34.
128. Alcon-LePoder S, Drouet MT, Roux P, Frenkiel MP, Arborio M, Durand-Schneider AM, et al. The secreted form of dengue virus nonstructural protein NS1 is endocytosed by hepatocytes and accumulates in late endosomes: implications for viral infectivity. J Virol. 2005;79(17):11403-11.
129. Muller DA, Young PR. The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. Antiviral Res. 2013;98(2):192-208.
130. Lin CF, Lei HY, Shiau AL, Liu CC, Liu HS, Yeh TM, et al. Antibodies from dengue patient sera cross-react with endothelial cells and induce damage. J Med Virol. 2003;69(1):82-90.
131. Youn S, Ambrose RL, Mackenzie JM, Diamond MS. Non-structural protein-1 is required for West Nile virus replication complex formation and viral RNA synthesis. Virol J. 2013;10:339.
132. Wilson JR, de Sessions PF, Leon MA, Scholle F. West Nile virus nonstructural protein 1 inhibits TLR3 signal transduction. J Virol. 2008;82(17):8262-71.
133. Yoon KJ, Song G, Qian X, Pan J, Xu D, Rho HS, et al. Zika-Virus-Encoded NS2A Disrupts Mammalian Cortical Neurogenesis by Degrading Adherens Junction Proteins. Cell Stem Cell. 2017;21(3):349-58.e6.
134. Dalrymple NA, Cimica V, Mackow ER. Dengue Virus NS Proteins Inhibit RIG-I/MAVS Signaling by Blocking TBK1/IRF3 Phosphorylation: Dengue Virus Serotype 1 NS4A Is a Unique Interferon-Regulating Virulence Determinant. MBio. 2015;6(3):e00553-15.
135. Liu WJ, Wang XJ, Clark DC, Lobigs M, Hall RA, Khromykh AA. A single amino acid substitution in the West Nile virus nonstructural protein NS2A disables its ability to inhibit alpha/beta interferon induction and attenuates virus virulence in mice. J Virol. 2006;80(5):2396-404.
136. Liu WJ, Chen HB, Wang XJ, Huang H, Khromykh AA. Analysis of adaptive mutations in Kunjin virus replicon RNA reveals a novel role for the flavivirus nonstructural protein NS2A in inhibition of beta interferon promoter-driven transcription. J Virol. 2004;78(22):12225-35.
137. Mackenzie JM, Khromykh AA, Jones MK, Westaway EG. Subcellular localization and some biochemical properties of the flavivirus Kunjin nonstructural proteins NS2A and NS4A. Virology. 1998;245(2):203-15.
138. Leung JY, Pijlman GP, Kondratieva N, Hyde J, Mackenzie JM, Khromykh AA. Role of Nonstructural Protein NS2A in Flavivirus Assembly. J Virol. 2008;82(10):4731-41.
139. Luo D, Xu T, Hunke C, Gruber G, Vasudevan SG, Lescar J. Crystal structure of the NS3 protease-helicase from dengue virus. J Virol. 2008;82(1):173-83.

140. Luo D, Vasudevan SG, Lescar J. The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. Antiviral Res. 2015;118:148-58.
141. Shiryaev SA, Farhy C, Pinto A, Huang CT, Simonetti N, Ngono AE, et al. Characterization of the Zika virus two-component NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists. Antiviral Res. 2017;143:218-29.
142. Arias CF, Preugschat F, Strauss JH. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. Virology. 1993;193(2):888-99.
143. Breton ML, Meyniel-Schicklin L, Deloire A, Coutard B, Canard B, Lamballerie Xd, et al. Flavivirus NS3 and NS5 proteins interaction network: a high-throughput yeast two-hybrid screen. BMC Microbiology. 2011;11(1):234.
144. Aguirre S, Maestre AM, Pagni S, Patel JR, Savage T, Gutman D, et al. DENV inhibits type I IFN production in infected cells by cleaving human STING. PLoS Pathog. 2012;8(10):e1002934.
145. Lin CW, Lin KH, Lyu PC, Chen WJ. Japanese encephalitis virus NS2B-NS3 protease binding to phage-displayed human brain proteins with the domain of trypsin inhibitor and basic region leucine zipper. Virus Res. 2006;116(1-2):106-13.
146. Heaton NS, Perera R, Berger KL, Khadka S, LaCount DJ, Kuhn RJ, et al. Dengue virus nonstructural protein 3 redistributes fatty acid synthase to sites of viral replication and increases cellular fatty acid synthesis. Proc Natl Acad Sci U S A. 2010;107(40):17345-50.
147. Zou J, Xie X, Wang QY, Dong H, Lee MY, Kang C, et al. Characterization of Dengue Virus NS4A and NS4B Protein Interaction. J Virol. 892015. p. 3455-70.
148. Lee CM, Xie X, Zou J, Li SH, Lee MY, Dong H, et al. Determinants of Dengue Virus NS4A Protein Oligomerization. J Virol. 2015;89(12):6171-83.
149. Miller S, Kastner S, Krijnse-Locker J, Buhler S, Bartenschlager R. The non-structural protein 4A of dengue virus is an integral membrane protein inducing membrane alterations in a 2K-regulated manner. J Biol Chem. 2007;282(12):8873-82.
150. Roosendaal J, Westaway EG, Khromykh A, Mackenzie JM. Regulated cleavages at the West Nile virus NS4A-2K-NS4B junctions play a major role in rearranging cytoplasmic membranes and Golgi trafficking of the NS4A protein. J Virol. 2006;80(9):4623-32.
151. McLean JE, Wudzinska A, Datan E, Quaglino D, Zakeri Z. Flavivirus NS4A-induced autophagy protects cells against death and enhances virus replication. J Biol Chem. 2011;286(25):22147-59.
152. Lee CJ, Liao CL, Lin YL. Flavivirus Activates Phosphatidylinositol 3-Kinase Signaling To Block Caspase-Dependent Apoptotic Cell Death at the Early Stage of Virus Infection. J Virol. 2005;79(13):8388-99.
153. Muñoz-Jordán JL, Sánchez-Burgos GG, Laurent-Rolle M, García-Sastre A. Inhibition of interferon signaling by dengue virus. Proc Natl Acad Sci U S A. 2003;100(24):14333-8.
154. Nemésio H, Palomares-Jerez F, Villalaín J. NS4A and NS4B proteins from dengue virus: Membranotropic regions. 2012;1818(11):2818–30.
155. Zou J, Xie X, Lee lT, Chandrasekaran R, Reynaud A, Yap L, et al. Dimerization of flavivirus NS4B protein. J Virol. 2014;88(6):3379-91.
156. Zou J, Lee le T, Wang QY, Xie X, Lu S, Yau YH, et al. Mapping the Interactions between the NS4B and NS3 proteins of dengue virus. J Virol. 2015;89(7):3471-83.
157. Grant D, Tan GK, Qing M, Ng JK, Yip A, Zou G, et al. A single amino acid in nonstructural protein NS4B confers virulence to dengue virus in AG129 mice through enhancement of viral RNA synthesis. J Virol. 2011;85(15):7775-87.
158. Ishikawa H, Ma Z, Barber GN. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. Nature. 2009;461(7265):788-92.
159. Hannemann H, Sung PY, Chiu HC, Yousuf A, Bird J, Lim SP, et al. Serotype-specific Differences in Dengue Virus Non-structural Protein 5 Nuclear Localization*. J Biol Chem. 2013;288(31):22621-35.
160. Kumar A, Buhler S, Selisko B, Davidson A, Mulder K, Canard B, et al. Nuclear localization of dengue virus nonstructural protein 5 does not strictly correlate with efficient viral RNA replication and inhibition of type I interferon signaling. J Virol. 2013;87(8):4545-57.
161. Tay MY, Smith K, Ng IH, Chan KW, Zhao Y, Ooi EE, et al. The C-terminal 18 Amino Acid Region of Dengue Virus NS5 Regulates its Subcellular Localization and Contains a Conserved Arginine Residue Essential for Infectious Virus Production. PLoS Pathog. 2016;12(9):e1005886.

162. Uchil PD, Kumar AV, Satchidanandam V. Nuclear localization of flavivirus RNA synthesis in infected cells. J Virol. 2006;80(11):5451-64.
163. Cui T, Sugrue RJ, Xu Q, Lee AK, Chan YC, Fu J. Recombinant dengue virus type 1 NS3 protein exhibits specific viral RNA binding and NTPase activity regulated by the NS5 protein. Virology. 1998;246(2):409-17.
164. Potisopon S, Ferron F, Fattorini V, Selisko B, B C. Substrate selectivity of Dengue and Zika virus NS5 polymerase towards 2'-modified nucleotide analogues. Elsevier. 2017;140:25–36.
165. Ray D, Shah A, Tilgner M, Guo Y, Zhao Y, Dong H, et al. West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. J Virol. 2006;80(17):8362-70.
166. Chang DC, Hoang LT, Mohamed Naim AN, Dong H, Schreiber MJ, Hibberd ML, et al. Evasion of early innate immune response by 2'-O-methylation of dengue genomic RNA. Virology. 2016;499:259-66.
167. Lichinchi G, Zhao BS, Wu Y, Lu Z, Qin Y, He C, et al. Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. Cell Host Microbe. 2016;20(5):666-73.
168. Morrison J, Laurent-Rolle M, Maestre AM, Rajsbaum R, Pisanelli G, Simon V, et al. Dengue virus co-opts UBR4 to degrade STAT2 and antagonize type I interferon signaling. PLoS Pathog. 2013;9(3):e1003265.
169. Chen S, Wu Z, Wang M, Cheng A. Innate Immune Evasion Mediated by Flaviviridae Non-Structural Proteins. Viruses. 2017;9(10).
170. Mairiang D, Zhang H, Sodja A, Murali T, Suriyaphol P, Malasit P, et al. Identification of new protein interactions between dengue fever virus and its hosts, human and mosquito. PLoS One. 2013;8(1):e53535.
171. De Maio FA, Risso G, Iglesias NG, Shah P, Pozzi B, Gebhard LG, et al. The Dengue Virus NS5 Protein Intrudes in the Cellular Spliceosome and Modulates Splicing. PLoS Pathog. 2016;12(8):e1005841.
172. Qing M, Yang F, Zhang B, Zou G, Robida JM, Yuan Z, et al. Cyclosporine inhibits flavivirus replication through blocking the interaction between host cyclophilins and viral NS5 protein. Antimicrob Agents Chemother. 2009;53(8):3226-35.
173. Baez CF, Barel VA, de Souza AM, Rodrigues CR, Varella RB, Cirauqui N. Analysis of worldwide sequence mutations in Zika virus proteins E, NS1, NS3 and NS5 from a structural point of view. Mol Biosyst. 2016;13(1):122-31.
174. Stephen P, Baz M, Boivin G, Lin SX. Structural Insight into NS5 of Zika Virus Leading to the Discovery of MTase Inhibitors. J Am Chem Soc. 2016;138(50):16212-5.
175. Coutard B, Barral K, Lichiere J, Selisko B, Martin B, Aouadi W, et al. Zika Virus Methyltransferase: Structure and Functions for Drug Design Perspectives. J Virol. 2017;91(5).
176. Rajapakse S, Rodrigo C, Rajapakse A. Treatment of dengue fever. Infect Drug Resist. 2012;5(2012:5):103-12.
177. Leyssen P, De Clercq E, Neyts J. Perspectives for the Treatment of Infections with Flaviviridae. Clin Microbiol Rev. 2000;13(1):67-82.
178. Sampath A, Padmanabhan R. Molecular targets for flavivirus drug discovery. Antiviral Res. 2009;81(1):6-15.
179. Zhang XG, Mason PW, Dubovi EJ, Xu X, Bourne N, Renshaw RW, et al. Antiviral activity of geneticin against dengue virus. Antiviral Res. 2009;83(1):21-7.
180. Reznik SE, Ashby CR, Jr. Sofosbuvir: an antiviral drug with potential efficacy against Zika infection. Int J Infect Dis. 2017;55:29-30.
181. Kochhar DM, Penner JD, Knudsen TB. Embryotoxic, teratogenic, and metabolic effects of ribavirin in mice. Toxicol Appl Pharmacol. 1980;52(1):99-112.
182. Chaudhuri G, Roy S, Das R, Arora RB. Interruption of pregnancy in rats by 6-azauridine. Indian J Med Res. 1974;62(12):1888-92.
183. Ihara T, Oneda S, Yamamoto T, Boudrel L, Lau D, Miller D, et al. An Embryotoxic/Teratogenic Potential and Abortifacient Effect Study of Interferon alfacon-1 (Infergen®) via Subcutaneous Administration to Rhesus Monkeys. Congenital Anomalies. 2017;39(4):223-42.
184. Ishikawa T, Yamanaka A, Konishi E. A review of successful flavivirus vaccines and the problems with those flaviviruses for which vaccines are not yet available. Vaccine. 2014;32(12):1326-37.
185. Angelo MA, Grifoni A, O'Rourke PH, Sidney J, Paul S, Peters B, et al. Human CD4+ T Cell Responses to an Attenuated Tetravalent Dengue Vaccine Parallel Those Induced by Natural Infection in Magnitude, HLA Restriction, and Antigen Specificity. J Virol. 2017;91(5).

186. Abbasi J. First Inactivated Zika Vaccine Trial. Jama. 2016;316(24):2588.
187. Maxian O, Neufeld A, Talis EJ, Childs LM, Blackwood JC. Zika virus dynamics: When does sexual transmission matter? Epidemics. 2017;21:48-55.
188. Brownstein JS, Hett E, O'Neill SL. The potential of virulent Wolbachia to modulate disease transmission by insects. J Invertebr Pathol. 2003;84(1):24-9.
189. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, et al. A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. Cell. 2009;139(7):1268-78.
190. Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, Muzzi F, et al. Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. Nature. 2011;476(7361):454-7.
191. Weaver SC. Urbanization and geographic expansion of zoonotic arboviral diseases: mechanisms and potential strategies for prevention. Trends Microbiol. 2013.
192. Rao PK. CCR5 inhibitors: Emerging promising HIV therapeutic strategy. Indian J Sex Transm Dis. 2009;30(1):1-9.
193. Krishnan MN, Garcia-Blanco MA. Targeting host factors to treat West Nile and dengue viral infections. Viruses. 2014;6(2):683-708.
194. Wagstaff K, Sivakumaran H, Heaton S, Harrich D, Jans D. Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. Biochem J. 2012;443(Pt 3):851-6.
195. Shoji-Kawata S, Sumpter R, Leveno M, Campbell GR, Zou Z, Kinch L, et al. Identification of a candidate therapeutic autophagy–inducing peptide. Nature. 2013;494(7436):201-6.
196. Verma S, Kumar M, Gurjav U, Lum S, Nerurkar VR. Reversal of West Nile virus-induced blood-brain barrier disruption and tight junction proteins degradation by matrix metalloproteinases inhibitor. Virology. 2010;397(1):130-8.
197. Souza DG, Fagundes CT, Sousa LP, Amaral FA, Souza RS, Souza AL, et al. Essential role of platelet-activating factor receptor in the pathogenesis of Dengue virus infection. Proc Natl Acad Sci U S A. 2009;106(33):14138-43.
198. Xing S, Wallmeroth N, Berendzen KW, Grefen C. Techniques for the Analysis of Protein-Protein Interactions in Vivo. Plant Physiol. 2016;171(2):727-58.
199. Wu Y, Li Q, Chen XZ. Detecting protein-protein interactions by Far western blotting. Nat Protoc. 2007;2(12):3278-84.
200. Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, et al. Proteomic Mapping of Mitochondria in Living Cells via Spatially-Restricted Enzymatic Tagging. Science. 2013;339(6125):1328-31.
201. Beck DB, Narendra V, Drury WJ, 3rd, Casey R, Jansen PW, Yuan ZF, et al. In vivo proximity labeling for the detection of protein-protein and protein-RNA interactions. J Proteome Res. 2014;13(12):6135-43.
202. Kotlyar M, Pastrello C, Pivetta F, Lo Sardo A, Cumbaa C, Li H, et al. In silico prediction of physical protein interactions and characterization of interactome orphans. Nat Methods. 2015;12(1):79-84.
203. Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, et al. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Mol Cell. 2016;61(3):341-51.
204. Firat-Karalar EN, Stearns T. Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. Methods Cell Biol. 2015;129:153-70.
205. Ritchie C, Cylinder I, Platt EJ, Barklis E. Analysis of HIV-1 Gag Protein Interactions via Biotin Ligase Tagging. J Virol. 2015;89(7):3988-4001.
206. Schmidt A, Forne I, Imhof A. Bioinformatic analysis of proteomics data. BMC Syst Biol. 2014;8(Suppl 2):S3.
207. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11(12):2301-19.
208. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731-40.
209. Zhang F, Hammack C, Ogden SC, Cheng Y, Lee EM, Wen Z, et al. Molecular signatures associated with ZIKV exposure in human cortical neural progenitors. Nucleic Acids Res. 2016;44(18):8610-20.
210. Yuan X, Shan Y, Zhao Z, Chen J, Cong Y. G0/G1 arrest and apoptosis induced by SARS-CoV 3b protein in transfected cells. Virol J. 2005;2:66.

211. Castedo M, Perfettini JL, Roumier T, Kroemer G. Cyclin-dependent kinase-1: linking apoptosis to cell cycle and mitotic catastrophe. Cell Death Differ. 2002;9(12):1287-93.
212. Doolittle JM, Gomez SM. Mapping protein interactions between Dengue virus and its human and insect hosts. PLoS Negl Trop Dis. 2011;5(2):e954.
213. Esteves E, Rosa N, Correia MJ, Arrais JP, Barros M. New Targets for Zika Virus Determined by Human-Viral Interactomic: A Bioinformatics Approach. Biomed Res Int. 2017;2017.
214. Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, et al. A physical interaction network of dengue virus and human proteins. Mol Cell Proteomics. 2011;10(12):M111.012187.
215. Poyomtip T, Hodge K, Matangkasombut P, Sakuntabhai A, Pisitkun T, Jirawatnotai S, et al. Development of viable TAP-tagged dengue virus for investigation of host-virus interactions in viral replication. J Gen Virol. 2016;97(3):646-58.
216. Li G, Poulsen M, Fenyvuesvolgyi C, Yashiroda Y, Yoshida M, Simard JM, et al. Characterization of cytopathic factors through genome-wide analysis of the Zika viral proteins in fission yeast. Proc Natl Acad Sci U S A. 2017;114(3):E376-e85.
217. Wang C, Yang SNY, Smith K, Forwood JK, Jans DA. Nuclear import inhibitor N-(4-hydroxyphenyl) retinamide targets Zika virus (ZIKV) nonstructural protein 5 to inhibit ZIKV infection. Biochem Biophys Res Commun. 2017;493(4):1555-9.
218. Coyaud E. RC, Cheng D., Gonçalves J., Dyakov B.J.A, Laurent E.M.N, St-Germain J., Pelletier L., Gingras A., Brumell J.H., Kim P.K., Safronetz D., Raught B. Global interactomics uncovers extensive organellar targeting by Zika virus. - PubMed - NCBI. Mol Cell Proteomics. 2018.
219. Reyes-Del Valle J, Chavez-Salinas S, Medina F, Del Angel RM. Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells. J Virol. 2005;79(8):4557-67.
220. Srisutthisamphan K, Jirakanwisal K, Ramphan S, Tongluan N, Kuadkitkan A, Smith DR. Hsp90 interacts with multiple dengue virus 2 proteins. Scientific Reports. 2018;8(1):4308.
221. Edgil D, Polacek C, Harris E. Dengue virus utilizes a novel strategy for translation initiation when cap-dependent translation is inhibited. J Virol. 2006;80(6):2976-86.
222. Porter FW, Bochkov YA, Albee AJ, Wiese C, Palmenberg AC. A picornavirus protein interacts with Ran-GTPase and disrupts nucleocytoplasmic transport. Proc Natl Acad Sci U S A. 2006;103(33):12417-22.
223. Aguirre S, Fernandez-Sesma A. Collateral Damage during Dengue Virus Infection: Making Sense of DNA by cGAS. J Virol. 2017;91(14).
224. Viktorovskaya OV, Greco TM, Cristea IM, Thompson SR. Identification of RNA Binding Proteins Associated with Dengue Virus RNA in Infected Cells Reveals Temporally Distinct Host Factor Requirements. PLoS Negl Trop Dis. 2016;10(8):e0004921.
225. Olagnier D, Peri S, Steel C, van Montfoort N, Chiang C, Beljanski V, et al. Cellular Oxidative Stress Response Controls the Antiviral and Apoptotic Programs in Dengue Virus-Infected Dendritic Cells. PLoS Pathog. 2014;10(12).
226. Rahman MM, Bagdassarian E, Ali MAM, McFadden G. Identification of host DEAD-box RNA helicases that regulate cellular tropism of oncolytic Myxoma virus in human cancer cells. Scientific Reports. 2017;7(1):15710.
227. Bortz E, Westera L, Maamary J, Steel J, Albrecht RA, Manicassamy B, et al. Host- and Strain-Specific Regulation of Influenza Virus Polymerase Activity by Interacting Cellular Proteins. mBio. 2011;2(4).
228. Cheng W, Chen G, Jia H, He X, Jing Z. DDX5 RNA Helicases: Emerging Roles in Viral Infection. Int J Mol Sci. 2018;19(4).
229. Taguwa S, Maringer K, Li X, Bernal-Rubio D, Rauch JN, Gestwicki JE, et al. Defining Hsp70 Subnetworks in Dengue Virus Replication Reveals Key Vulnerability in Flavivirus Infection. Cell. 2015;163(5):1108-23.

6: Appendix

6.1: Appendix A – Recipes

6.1.1: Cell culture

Growth media (GM):

- Dulbecco's modified eagle's medium (DMEM) + Glutamax
- 10% FBS
- 0.1mM non-essential amino acids

Luria broth (LB):

- 10g bacteriological peptone
- 5g yeast extract
- 5g NaCl
- Up to 1L ddH_2O and autoclave

LB agar:

- 10g bacteriological peptone
- 5g yeast extract
- 5g NaCl
- 1g agar powder
- Up to 1L ddH_2O and autoclave

6.1.2: DNA/RNA techniques

10x TBE buffer:

- 108g Tris base
- 55g boric acid
- 40ml 0.5M EDTA
- Up to 1L with ddH_2O and autoclave

1% agarose gel:

- 1g agarose powder
- 100ml 1xTBE

6.1.3: Protein techniques

6x sample buffer:

- 1.2g SDS
- 7ml 0.5M Tris-HCl pH6.8
- 3ml glycerol

0.5M Tris-HCl pH6.8:

- 6g Tris base
- 100ml ddH_2O
- Adjust to pH6.8 with conc. HCl

1.5M Tris-HCl pH 8.8:

- 36.3g Tris base
- 200ml ddH_2O
- Adjust to pH8.8 with conc. HCl

4% paraformaldehyde:

- 40g paraformaldehyde powder
- 1L PBS
- Dissolve using conc. NaOH and heat
- Adjust to pH6.9 with conc. HCl

Coomassie Brilliant Blue dye:

- 0.575g Coomassie Brilliant Blue R-250 powder
- 200ml methanol

- 50ml acetic acid
- 250ml ddH_2O

Coomassie destain solution:

- 50ml glacial acetic acid
- 350ml ddH_2O
- 100ml methanol

10% resolving SDS PAGE gel:

- 2.3ml ddH_2O
- 1.3ml 40% polyacrylamide
- 1.3ml 1.5M Tris-HCl pH8.8
- 50µl 10% SDS
- 50µl 10% APS
- 3µl TEMED

5% stacking SDS PAGE gel:

- 2.4ml ddH_2O
- 0.5 ml 40% polyacrylamide
- 1ml 0.5M Tris-HCl pH6.8
- 40µl 10% SDS
- 40µl 10% APS
- 4µl TEMED

10x SDS PAGE running buffer:

- 30.275g Tris base
- 144g glycine
- 50ml 20% SDS
- Up to 1L with ddH_2O

1x transfer buffer:

- 100ml methanol

- 12.5g Tris base
- 5.63g glycine
- Up to 500ml with ddH_2O
- Adjust to pH8.3 with acetic acid

Western blot blocking solution:

- 2g skim milk powder
- 40ml 0.1% tween (PBS)

Streptavidin western blot blocking solution:

- 5% BSA in 0.1% tween (PBS)

10x TBS:

- 12.125g Trizma HCl
- 40.0g NaCl
- Up to 500ml ddH_2O
- Adjust to pH7.6 with conc. HCl

Pulldown lysis buffer:

- 1ml 10xTBS
- 10µl 500mM EDTA
- 1ml 10% triton X-100 (PBS)
- 6.6ml ddH_2O
- 7x protease inhibitor tablets (in 1.4ml ddH_2O)

Pulldown equilibration buffer:

- 1ml 10xTBS
- 9ml ddH_2O

Pulldown wash buffer:

- 1ml 10x TBS
- 10µl 500mM EDTA
- 0.25ml 10% Triton X-100 (PBS)
- 8.72ml ddH_2O

Streptavidin pulldown Wash buffer 1:

- 2% SDS (ddH_2O)

Streptavidin pulldown Wash buffer 2:

- 0.2% Sodium deoxycholate
- 1% Triton X-100
- 500mM NaCl
- 1mM EDTA
- 50mM HEPES pH7.5

Streptavidin pulldown Wash buffer 3:

- 10mM Tris pH8.1
- 250mM LiCl
- 0.5% Sodium deoxycholate
- 1% Triton X-100
- 500mM NaCl
- 1mM EDTA
- 0.5% NP-40

Streptavidin pulldown Wash buffer 4:

- 50mM Tris pH7.4
- 50mM NaCl

LC Buffer A:

- 20mM ammonium hydroxide pH10

LC Buffer B:

- 20mM ammonium hydroxide (acetonitrile) pH10

6.2: Appendix B – Sequences

6.2.1: ZNS5-FLAG gene

ATGGGCGGAGGCACAGGCGAGACACTGGGCGAGAAGTGGAAGGCCCGGCTGAACCAGATG
AGCGCCCTGGAGTTCTACAGCTACAAGAAGTCCGGCATCACCGAAGTGTGCCGGGAAGAGG
CCAGACGGGCCCTGAAGGATGGCGTGGCAACAGGCGGCCACGCCGTGTCTAGAGGCAGCG
CCAAGCTGCGGTGGCTGGTGGAAAGAGGCTACCTGCAGCCCTACGGCAAAGTGATCGACCT
GGGCTGCGGCAGAGGCGGCTGGTCTTACTACGCCGCCACCATCCGGAAGGTGCAGGAAGTG
AAGGGCTACACCAAGGGCGGACCCGGCCACGAGGAACCCGTGCTGGTGCAGAGCTACGGC
TGGAACATCGTGCGGCTGAAGTCCGGCGTGGACGTGTTCCACATGGCCGCCGAGCCCTGCG
ACACCCTGCTGTGCGATATCGGCGAGAGCAGCAGCAGCCCCGAGGTGGAAGAGGCCCGGA
CCCTGAGAGTGCTGAGCATGGTCGGAGACTGGCTGGAAAAGCGGCCTGGCGCCTTCTGCAT
CAAGGTGCTGTGCCCCTACACCAGCACCATGATGGAAACCCTGGAACGGCTGCAGCGGAGA
TACGGCGGAGGACTCGTGCGGGTGCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGG
TGTCCGGCGCCAAGAGCAACACCATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGG
CAGAATGGACGGCCCCAGACGGCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGC
ACCAGAGCCGTGGTGTCCTGCGCCGAGGCCCCCAACATGAAGATCATCGGCAACCGGATCG
AGCGGATCAGAAGCGAGCACGCCGAGACATGGTTCTTCGACGAGAACCACCCCTACCGGAC
CTGGGCCTACCACGGCAGCTACGAGGCCCCTACACAGGGCAGCGCCAGCAGCCTGATCAAC
GGCGTCGTCCGGCTGCTGAGCAAGCCCTGGGATGTGGTCACCGGCGTGACCGGAATCGCCA
TGACCGACACCACCCCCTACGGCCAGCAGCGGGTGTTCAAAGAAAAGGTGGACACCCGGGT
GCCCGACCCCCAGGAAGGCACCAGACAGGTCATGAGCATGGTGTCCAGCTGGCTCTGGAAA
GAGCTGGGCAAGCACAAGCGGCCCAGAGTGTGCACCAAAGAAGAGTTCATCAACAAAGTGC
GGAGCAACGCCGCCCTGGGCGCCATCTTCGAGGAAGAGAAAGAGTGGAAAACCGCCGTCGA
GGCCGTGAACGACCCCCGGTTTTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACCTGAGA
GGCGAGTGCCAGAGCTGCGTGTACAACATGATGGGCAAGCGCGAGAAGAAGCAGGGCGAGT
TCGGCAAGGCCAAGGGCAGCCGGGCCATCTGGTACATGTGGCTGGGCGCCAGATTCCTGGA
ATTTGAGGCCCTGGGCTTCCTGAACGAGGACCACTGGATGGGCAGAGAGAACAGCGGCGGA
GGCGTGGAAGGCCTGGGCCTGCAGAGACTGGGCTACGTGCTGGAAGAGATGAGCAGAATCC
CTGGCGGCAGGATGTACGCCGACGACACCGCCGGCTGGGACACCCGGATCAGCAGATTCGA
CCTGGAAAACGAGGCCCTGATCACCAACCAGATGGAAAAGGGCCACCGGGCCCTGGCCCTG
GCCATCATCAAGTACACCTACCAGAACAAGGTGGTCAAAGTGCTGCGGCCTGCCGAGAAGGG
CAAGACCGTGATGGACATCATCAGCCGGCAGGACCAGCGGGGCAGCGGCCAGGTGGTCAC
CTACGCCCTGAACACCTTCACCAACCTGGTGGTGCAGCTGATCCGGAACATGGAAGCCGAAG
AGGTGCTGGAAATGCAGGACCTGTGGCTGCTGCGGCGGAGCGAGAAAGTGACCAACTGGCT
GCAGTCCAACGGCTGGGACCGGCTGAAGAGAATGGCCGTGTCCGGCGACGACTGCGTGGTC
AAGCCCATCGACGACAGATTCGCCCACGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCG
GAAGGACACCCAGGAATGGAAGCCCAGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTTC

TGCAGCCACCACTTCAACAAGCTGCACCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCA
GACACCAGGACGAGCTGATCGGCAGAGCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAG
AGAGACAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGG
CGGGACCTGCGGCTGATGGCCAACGCCATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCA
CCGGCAGAACCACCTGGTCCATCCACGGCAAGGGCGAGTGGATGACCACCGAGGACATGCT
GGTCGTGTGGAACAGAGTGTGGATCGAAGAGAACGACCACATGGAAGATAAGACCCCCGTG
ACCAAGTGGACCGACATCCCCTACCTGGGCAAGCGGGAAGATCTGTGGTGCGGCTCCCTGA
TCGGCCACAGACCTAGGACCACCTGGGCCGAGAACATCAAGAACACCGTGAACATGGTCCG
ACGGATCATCGGCGACGAAGAGAAGTACATGGACTACCTGAGCACACAAGTGCGATACCTGG
GCGAGGAAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAGACTACAAGGACCA
CGACGGTGACTACAAGGACCACGACATCGACTACAAGGACGACGACGACAAG

Red = FLAG tag

Orange = RE site

6.2.2: ZNS5-BirA gene with sequencing primers

ATGGGCGGAGGCACAGGCGAGACACTGGGCGAGAAGTGGAAGGCCCGGCTGAACCAGATG
AGCGCCCTGGAGTTCTACAGCTACAAGAAGTCCGGCATCACCGAAGTGTGCCGGGAAGAGG
CCAGACGGGCCCTGAAGGATGGCGTGGCAACAGGCGGCCACGCCGTGTCTAGAGGCAGCG
CCAAGCTGCGGTGGCTGGTGGAAAGAGGCTACCTGCAGCCCTACGGCAAAGTGATCGACCT
GGGCTGCGGCAGAGGCGGCTGGTCTTACTACGCCGCCACCATCCGGAAGGTGCAGGAAGTG
AAGGGCTACACCAAGGGCGGACCCGGCCACGAGGAACCCGTGCTGGTGCAGAGCTACGGC
TGGAACATCGTGCGGCTGAAGTCCGGCGTGGACGTGTTCCACATGGCCGCCGAGCCCTGCG
ACACCCTGCTGTGCGATATCGGCGAGAGCAGCAGCAGCCCCGAGGTGGAAGAGGCCCGGA
CCCTGAGAGTGCTGAGCATGGTCGGAGACTGGCTGGAAAAGCGGCCTGGCGCCTTCTGCAT
CAAGGTGCTGTGCCCCTACACCAGCACCATGATGGAAACCCTGGAACGGCTGCAGCGGAGA
TACGGCGGAGGACTCGTGCGGGTGCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGG
TGTCCGGCGCCAAGAGCAACACCATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGG
CAGAATGGACGGCCCCAGACGGCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGC
ACCAGAGCCGTGGTGTCCTGCGCCGAGGCCCCCAACATGAAGATCATCGGCAACCGGATCG
AGCGGATCAGAAGCGAGCACGCCGAGACATGGTTCTTCGACGAGAACCACCCCTACCGGAC
CTGGGCCTACCACGGCAGCTACGAGGCCCCTACACAGGGCAGCGCCAGCAGCCTGATCAAC
GGCGTCGTCCGGCTGCTGAGCAAGCCCTGGGATGTGGTCACCGGCGTGACCGGAATCGCCA
TGACCGACACCACCCCCTACGGCCAGCAGCGGGTGTTCAAAGAAAAGGTGGACACCCGGGT
GCCCGACCCCCAGGAAGGCACCAGACAGGTCATGAGCATGGTGTCCAGCTGGCTCTGGAAA
GAGCTGGGCAAGCACAAGCGGCCCAGAGTGTGCACCAAAGAAGAGTTCATCAACAAAGTGC
GGAGCAACGCCGCCCTGGGCGCCATCTTCGAGGAAGAGAAAGAGTGGAAAACCGCCGTCGA
GGCCGTGAACGACCCCCGGTTTTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACCTGAGA
GGCGAGTGCCAGAGCTGCGTGTACAACATGATGGGCAAGCGCGAGAAGAAGCAGGGCGAGT
TCGGCAAGGCCAAGGGCAGCCGGGCCATCTGGTACATGTGGCTGGGCGCCAGATTCCTGGA
ATTTGAGGCCCTGGGCTTCCTGAACGAGGACCACTGGATGGGCAGAGAGAACAGCGGCGGA
GGCGTGGAAGGCCTGGGCCTGCAGAGACTGGGCTACGTGCTGGAAGAGATGAGCAGAATCC
CTGGCGGCAGGATGTACGCCGACGACACCGCCGGCTGGGACACCCGGATCAGCAGATTCGA
CCTGGAAAACGAGGCCCTGATCACCAACCAGATGGAAAAGGGCCACCGGGCCCTGGCCCTG
GCCATCATCAAGTACACCTACCAGAACAAGGTGGTCAAAGTGCTGCGGCCTGCCGAGAAGGG
CAAGACCGTGATGGACATCATCAGCCGGCAGGACCAGCGGGGCAGCGGCCAGGTGGTCAC
CTACGCCCTGAACACCTTCACCAACCTGGTGGTGCAGCTGATCCGGAACATGGAAGCCGAAG
AGGTGCTGGAAATGCAGGACCTGTGGCTGCTGCGGCGGAGCGAGAAAGTGACCAACTGGCT
GCAGTCCAACGGCTGGGACCGGCTGAAGAGAATGGCCGTGTCCGGCGACGACTGCGTGGTC
AAGCCCATCGACGACAGATTCGCCCACGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCG
GAAGGACACCCAGGAATGGAAGCCCAGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTTC
TGCAGCCACCACTTCAACAAGCTGCACCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCA

GACACCAGGACGAGCTGATCGGCAGAGCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAG
AGAGACAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGG
CGGGACCTGCGGCTGATGGCCAACGCCATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCA
CCGGCAGAACCACCTGGTCCATCCACGGCAAGGGCGAGTGGATGACCACCGAGGACATGCT
GGTCGTGTGGAACAGAGTGTGGATCGAAGAGAACGACCACATGGAAGATAAGACCCCCGTG
ACCAAGTGGACCGACATCCCCTACCTGGGCAAGCGGGAAGATCTGTGGTGCGGCTCCCTGA
TCGGCCACAGACCTAGGACCACCTGGGCCGAGAACATCAAGAACACCGTGAACATGGTCCG
ACGGATCATCGGCGACGAAGAGAAGTACATGGACTACCTGAGCACACAAGTGCGATACCTGG
GCGAGGAAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAACACCGT
GCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGGCGAGCAGCTGGGA
GAGACCCTGGGAATGAGCAGAGCCGCCATCAACAAGCACATCCAGACACTGAGAGACTGGG
GAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCTGAGCCTATCCAGCTGCT
GAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAGCGTGGCCGTGCTGCCTGTGATC
GACTCCACCAATCAGTACCTGCTGGACAGAATCGGAGAGCTGAAGTCCGGCGACGCCTGCAT
CGCCGAGTACCAGCAGGCTGGCAGAGGAGGCAGAGGACGGAAGTGGTTCAGCCCATTCGGA
GCCAACCTGTACCTGTCCATGTTCTGGAGACTGGAGCAGGGACCTGCTGCTGCCATCGGACT
GAGTCTGGTGATCGGAATCGTGATGGCCGAGGTGCTGAGAAAGCTGGGAGCCGACAAGGTG
AGAGTGAAGTGGCCTAATGACCTGTACCTCCAGGACCGCAAGCTGGCTGGCATCCTGGTGGA
GCTGACAGGCAAGACAGGCGATGCCGCTCAGATCGTGATCGGAGCCGGAATCAACATGGCC
ATGAGAAGAGTGGAGGAGAGCGTGGTGAACCAGGGCTGGATCACCCTGCAGGAGGCTGGCA
TCAACCTGGACCGGAACACCCTGGCCGCCATGCTGATCAGAGAGCTGAGAGCCGCTCTGGA
GCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTGAGCAGATGGGAGAAGCTGGACAACTTCA
TCAACAGACCTGTGAAGCTGATCATCGGCGACAAGGAAATCTTCGGCATCTCCAGAGGAATC
GACAAGCAGGGAGCTCTGCTGCTGGAGCAGGACGGAATCATCAAGCCCTGGATGGGCGGAG
AAATCTCCCTGAGAAGCGCAGAGAAGTGA

Orange = RE site

Green = primer sites

Blue = BirA tag

6.2.3: ZNS5-BirA with RT-PCR primers

CGACTCACTATAGGGCGAATTCGGATCCGCCACCATGGGCGGAGGCACAGGCGAGACACTG
GGCGAGAAGTGGAAGGCCCGGCTGAACCAGATGAGCGCCCTGGAGTTCTACAGCTACAAGA
AGTCCGGCATCACCGAAGTGTGCCGGGAAGAGGCCAGACGGGCCCTGAAGGATGGCGTGG
CAACAGGCGGCCACGCCGTGTCTAGAGGCAGCGCCAAGCTGCGGTGGCTGGTGGAAAGAG
GCTACCTGCAGCCCTACGGCAAAGTGATCGACCTGGGCTGCGGCAGAGGCGGCTGGTCTTA
CTACGCCGCCACCATCCGGAAGGTGCAGGAAGTGAAGGGCTACACCAAGGGCGGACCCGG
CCACGAGGAACCCGTGCTGGTGCAGAGCTACGGCTGGAACATCGTGCGGCTGAAGTCCGGC
GTGGACGTGTTCCACATGGCCGCCGAGCCCTGCGACACCCTGCTGTGCGATATCGGCGAGA
GCAGCAGCAGCCCCGAGGTGGAAGAGGCCCGGACCCTGAGAGTGCTGAGCATGGTCGGAG
ACTGGCTGGAAAAGCGGCCTGGCGCCTTCTGCATCAAGGTGCTGTGCCCCTACACCAGCAC
CATGATGGAAACCCTGGAACGGCTGCAGCGGAGATACGGCGGAGGACTCGTGCGGGTGCCC
CTGAGCAGAAACAGCACCCACGAGATGTACTGGGTGTCCGGCGCCAAGAGCAACACCATCAA
GAGCGTGTCCACCACCAGCCAGCTGCTGCTGGGCAGAATGGACGGCCCCAGACGGCCCGT
GAAGTACGAAGAGGACGTGAACCTGGGCTCTGGCACCAGAGCCGTGGTGTCCTGCGCCGAG
GCCCCCAACATGAAGATCATCGGCAACCGGATCGAGCGGATCAGAAGCGAGCACGCCGAGA
CATGGTTCTTCGACGAGAACCACCCCTACCGGACCTGGGCCTACCACGGCAGCTACGAGGC
CCCTACACAGGGCAGCGCCAGCAGCCTGATCAACGGCGTCGTCCGGCTGCTGAGCAAGCCC
TGGGATGTGGTCACCGGCGTGACCGGAATCGCCATGACCGACACCACCCCCTACGGCCAGC
AGCGGGTGTTCAAAGAAAAGGTGGACACCCGGGTGCCCGACCCCCAGGAAGGCACCAGACA
GGTCATGAGCATGGTGTCCAGCTGGCTCTGGAAAGAGCTGGGCAAGCACAAGCGGCCCAGA
GTGTGCACCAAAGAAGAGTTCATCAACAAAGTGCGGAGCAACGCCGCCCTGGGCGCCATCTT
CGAGGAAGAGAAAGAGTGGAAAACCGCCGTCGAGGCCGTGAACGACCCCCGGTTTTGGGCC
CTGGTGGACAAAGAGAGAGAGCACCACCTGAGAGGCGAGTGCCAGAGCTGCGTGTACAACA
TGATGGGCAAGCGCGAGAAGAAGCAGGGCGAGTTCGGCAAGGCCAAGGGCAGCCGGGCCA
TCTGGTACATGTGGCTGGGCGCCAGATTCCTGGAATTTGAGGCCCTGGGCTTCCTGAACGAG
GACCACTGGATGGGCAGAGAGAACAGCGGCGGAGGCGTGGAAGGCCTGGGCCTGCAGAGA
CTGGGCTACGTGCTGGAAGAGATGAGCAGAATCCCTGGCGGCAGGATGTACGCCGACGACA
CCGCCGGCTGGGACACCCGGATCAGCAGATTCGACCTGGAAAACGAGGCCCTGATCACCAA
CCAGATGGAAAAGGGCCACCGGGCCCTGGCCCTGGCCATCATCAAGTACACCTACCAGAAC
AAGGTGGTCAAAGTGCTGCGGCCTGCCGAGAAGGGCAAGACCGTGATGGACATCATCAGCC
GGCAGGACCAGCGGGGCAGCGGCCAGGTGGTCACCTACGCCCTGAACACCTTCACCAACCT
GGTGGTGCAGCTGATCCGGAACATGGAAGCCGAAGAGGTGCTGGAAATGCAGGACCTGTGG
CTGCTGCGGCGGAGCGAGAAAGTGACCAACTGGCTGCAGTCCAACGGCTGGGACCGGCTGA
AGAGAATGGCCGTGTCCGGCGACGACTGCGTGGTCAAGCCCATCGACGACAGATTCGCCCA
CGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCGGAAGGACACCCAGGAATGGAAGCCC
AGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTTCTGCAGCCACCACTTCAACAAGCTGCA

CCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCAGACACCAGGACGAGCTGATCGGCAGA
GCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAGAGAGACAGCCTGCCTGGCCAAGAGCT
ACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGGCGGGACCTGCGGCTGATGGCCAACGC
CATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCACCGGCAGAACCACCTGGTCCATCCAC
GGCAAGGGCGAGTGGATGACCACCGAGGACATGCTGGTCGTGTGGAACAGAGTGTGGATCG
AAGAGAACGACCACATGGAAGATAAGACCCCCGTGACCAAGTGGACCGACATCCCCTACCTG
GGCAAGCGGGAAGATCTGTGGTGCGGCTCCCTGATCGGCCACAGACCTAGGACCACCTGGG
CCGAGAACATCAAGAACACCGTGAACATGGTCCGACGGATCATCGGCGACGAAGAGAAGTAC
ATGGACTACCTGAGCACACAAGTGCGATACCTGGGCGAGGAAGGCAGCACCCCCGGAGTGC
TGGCAGGAAAGCTTGCAGGAAAGGACAACACCGTGCCCCTGAAGCTGATCGCCCTGCTGGC
CAACGGCGAGTTCCACTCTGGCGAGCAGCTGGGAGAGACCCTGGGAATGAGCAGAGCCGCC
ATCAACAAGCACATCCAGACACTGAGAGACTGGGGAGTGGACGTGTTCACCGTGCCTGGCAA
GGGCTACAGCCTGCCTGAGCCTATCCAGCTGCTGAACGCCAAGCAGATCCTGGGACAGCTG
GATGGCGGAAGCGTGGCCGTGCTGCCTGTGATCGACTCCACCAATCAGTACCTGCTGGACA
GAATCGGAGAGCTGAAGTCCGGCGACGCCTGCATCGCCGAGTACCAGCAGGCTGGCAGAGG
AGGCAGAGGACGGAAGTGGTTCAGCCCATTCGGAGCCAACCTGTACCTGTCCATGTTCTGGA
GACTGGAGCAGGGACCTGCTGCTGCCATCGGACTGAGTCTGGTGATCGGAATCGTGATGGC
CGAGGTGCTGAGAAAGCTGGGAGCCGACAAGGTGAGAGTGAAGTGGCCTAATGACCTGTAC
CTCCAGGACCGCAAGCTGGCTGGCATCCTGGTGGAGCTGACAGGCAAGACAGGCGATGCCG
CTCAGATCGTGATCGGAGCCGGAATCAACATGGCCATGAGAAGAGTGGAGGAGAGCGTGGT
GAACCAGGGCTGGATCACCCTGCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCC
GCCATGCTGATCAGAGAGCTGAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTC
CTTACCTGAGCAGATGGGAGAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATC
GGCGACAAGGAAATCTTCGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGG
AGCAGGACGGAATCATCAAGCCCTGGATGGGCGGAGAAATCTCCCTGAGAAGCGCAGAGAA
GTGAC

Orange = RE sites

Green = Forward primer sites

Red = Reverse primer sites

Blue = BirA tag

6.2.4: ZNS5-BirA RT-PCR sequencing results

C3:

>prStart C3_prStart – Blast similarity = 97%

GTCCAGAACAGCAGCAGCCCCGAGGTGGAAGAGCCCGGACCCTGAGATGCTGAGATGGTC

GGAGACTGGCTGGAAAGCGGCCTGGCGCCTTCTGATCAAGGTGCTGTGCCCCTACACCAG

CACCATGATGGAAACCCTGGAACGGCTGCAGCGGAGATACGGCGGAGGACTCGTGCGGGT

GCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGGTGTCCGGCGCCAAGAGCAACAC

CATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGACAGAATGGACGGCCCCAGACG

GCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGCACCAGAGCCGTGTGTCCTGC

GCCGAGGCCCCAACATGAAGATCATCGGCACCGGACGAGCGGATCAGAACGAGCACGCCG

AGACATGGTCTTCGAC

>pr582 C3_pr582 – Blast similarity = 99%

GGAGACTCGTGCGGGTGCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGGTGTCCG

GCGCCAAGAGCAACACCATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGGCAGAA

TGGACGGCCCCAGACGGCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGCACCA

GAGCCGTGGTGTCCTGCGCCGAGGCCCCCAACATGAAGATCATCGGCAACCGGATCGAGC

GGATCAGAAGCGAGCACGCCGAGACATGGTTCTTCGACGAGAACCACCCCTACCGGACCT

GGGCCTACCACGGCAGCTACGAGGCCCCTACACAGGGCAGCGCCAGCAGCCTGATCAACG

GCGTCGTCCGGCTGCTGAGCAAGCCCTGGGATGTGGTCACCGGCGTGACCGGAATCGCCA

TGACCGACACCACCCCCTACGGCCAGCAGCGGGTGTTCAAAGAAAAGGTGGACACCCGGG

TGCCCGACCCCCAGGAAGGCACCAGACAGGTCATGAGCATGGTGTCCAGCTGGCTCTGGA

AAGAGCTGGGCAAGCACAAGCGGCCCAGAGTGTGCACCAAAGAAGAGTTCATCAACAAAG

TGCGGAGCAACGCCGCCCTGGGCGCCATCTTCGAGGAAGAGAAAGAGTGGAAAACCGCCG

TCGAGGCCGTGAACGACCCCCGGTTTTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACC

TGAGAGGCGAGTGCCAGAGCTGCGTGTACAACATGATGGGCAAGCGCGAGAAGAAGCAGG

GGCGAGTTCGGCAAGGCCAAGGGCAGCCGGGCCATCTGGTACATGTGGCTGGGCGCCAGA

TTCCTGGAATTTGAGGCCCTGGGCTTCCTGAACGAGGACCACTGGATGGGCAGAGAGAAC

AGCGGCGGAGGCGTGGAAGGCCTGGGCCTGCAGAGACTGGGCTACGTGCTGGAAGAGATG

AGCAGAATCCCTGGGCGGCAGGATGTACGCCGACGAC

>pr1260 C3_pr1260 – Blast similarity = 99%

AGCGCGAGAAGAAGCAGGGCGAGTTCGGCAAGGCCAAGGGCAGCCGGGCCATCTGGTACA

TGTGGCTGGGCGCCAGATTCCTGGAATTTGAGGCCCTGGGCTTCCTGAACGAGGACCACT

GGATGGGCAGAGAGAACAGCGGCGGAGGCGTGGAAGGCCTGGGCCTGCAGAGACTGGGCT

ACGTGCTGGAAGAGATGAGCAGAATCCCTGGCGGCAGGATGTACGCCGACGACACCGCCG

GCTGGGACACCCGGATCAGCAGATTCGACCTGGAAAACGAGGCCCTGATCACCAACCAGA

TGGAAAAGGGCCACCGGGCCCTGGCCCTGGCCATCATCAAGTACACCTACCAGAACAAGG

TGGTCAAAGTGCTGCGGCCTGCCGAGAAGGGCAAGACCGTGGATGGACATCATCAG

>prMid C3_prMid – Blast similarity = 99%

CATCCATCAGCCGGCAGGACCAGCGGGGCAGCGGCCAGGTGGTCACCTACGCCCTGAACA

CCTTCACCAACCTGGTGGTGCAGCAGGATCCGGAACATGGAAGCCGAAGAGGTGCTGGAA

ATGCAGGACCTGTGGCTGCTGCGGCGGAGCGAGAAAGTGACCAACTGGCTGCAGTCCAAC

GGCTGGGACCGGCTGAAGAGAATGGCCGTGTCCGGCGACGACTGCGTGGTCAAGCCCATC

GACGACAGATTCGCCCACGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCGGAAGGAC

ACCCAGGAATGGAAGCCCAGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTTCTGCAGC

CACCACTTCAACAAGCTGCACCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCAGACAC

CAGGACGAGCTGATCGGCAGAGCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAGAGAG

ACAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGGCGG

GACCTGCGGCTGATGGCCAACGCCATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCACC

GGCAGAACCACCTGGTCCATCCACGGCAAGGGCGAGTGGATGACCACCGAGGACATGCTG

GTCGTGTGGAACAGAGTGTGGATCGAAGAGAACGACCACATGGAAGATAAGACCCCCGTG

ACCAAGTGGACCGACATCCCCTACCTGGGCAAGCGGGAAGATCTGTGGTGCGGCTCCCTG

ATCGGCCACAGACCTAGGACCACCTGGGCCGAGAACATCAAGAACACCGTGAACATGGTC

CGACGGATCATCGGCGACGAAGAGAAGTACATGGACTACCTGAGCACACAAGTGCGATAC

CTGGGGCGAGGAAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAA

CACCGTGCCCCTGAAGCTG

>pr1980 C3_pr1980 – Blast similarity = 99%

CGCCATCGACGACAGATTCGCCCACGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCG

GAAGGACACCCAGGAATGGAAGCCCAGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTT

CTGCAGCCACCACTTCAACAAGCTGCACCTGAAGGACGGCCGGTCCATCGTGGTGCCCTG

CAGACACCAGGACGAGCTGATCGGCAGAGCCCGGGTGTCACCTGGCGCCGGATGGTCCAT

CAGAGAGACAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTGGCAGCTGCTGTACTTCCA

CCGGCGGGACCTGCGGCTGATGGCCAACGCCATCTGCAGCAGCGTGCCCGTGGATTGGGT

GCCCACCGGCAGAACCACCTGGTCCATCCACGGCAAGGGCGAGTGGATGACCACCGAGGA

CATGCTGGTCGTGTGGAACAGAGTGTGGATCGAAGAGAACGACCACATGGAAGATAAGAC

CCCCGTGACCAAGTGGACCGACATCCCCTACCTGGGCAAGCGGGAAGATCTGTGGTGCGG

CTCCCTGATCGGCCACAGACCTAGGACCACCTGGGCCGAGAACATCAAGAACACCGTGAA

CATGGTCCGACGGATCATCGGCGACGAAGAGAAGTACATGGACTACCTGAGCACACAAGT

GCGATACCTGGGCGAGGAAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAA

GGACAACACCGTGCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGG

CGAGCAGCTGGGAGAGACCCTGGGAATGACAAAAAGCCGCCATCAACAAGCACATCCAGA

CACTGAGAGACTGGGGAAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCT

GAGCCTATCCAGCTGCTGAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAA

>pr2647 C3_pr2647 – Blast similarity = 99%

GCGAGGAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAACACCGT

GCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGGCGAGCAGCTGGG

AGAGACCCTGGGAATGAGCAGAGCCGCCATCAACAAGCACATCCAGACACTGAGAGACTG

GGGAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCTGAGCCTATCCAGCT

GCTGAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAGCGTGGCCGTGCTGCCTGT

GATCGACTCCACCAATCAGTACCTGCTGGACAGAATCGGAGAGCTGAAGTCCGGCGACGC

CTGCATCGCCGAGTACCAGCAGGCTGGCAGAGGAGGCAGAGGACGGAAGTGGTTCAGCCC

ATTCGGAGCCAACCTGTACCTGTCCATGTTCTGGAGACTGGAGCAGGGACCTGCTGCTGC

CATCGGACTGAGTCTGGTGATCGGAATCGTGATGGCCGAGGTGCTGAGAAAGCTGGGAGC

CGACAAGGTGAGAGTGAAGTGGCCTAATGACCTGTACCTCCAGGACCGCAAGCTGGCTGG

CATCCTGGTGGAGCTGACAGGCAAGACAGGCGATGCCGCTCAGATCGTGATCGGAGCCGG

AATCAACATGGCCATGAGAAGAGTGGAGGAGAGCGTGGTGAACCAGGGCTGGATCACCCT

GCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCCGCCATGCTGATCAGAGAGCT

GAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTGAGCAGATGGGA

GAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATCGGCGACAAGGAAATCTT

CGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGGAGCAGGACGAATCTC

>prB621 C3_prB62 – Blast similarity = 99%

CTGGATCCCCTGCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCCGCCATGCTG

ATCAGAGAGCTGAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTG

AGCAGATGGGAGAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATCGGCGAC

AAGGAAATCTTCGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGGAGCAG

GACGGAATCATCAAGCCCTGGATGGGCGGAGAAATCTCCCTGAGAAGCGAAAAGA

C4:

>prStart C4_prStart – Blast similarity = 99%

GCCAGAACAGCAGCAGCCCCGAGGTGGAAGAGGCCCGGACCCTGAGAGTGCTGAGCATGG

TCGGAGACTGGCTGGAAAAGCGGCCTGGCGCCTTCTGCATCAAGGTGCTGTGCCCCTACA

CCAGCACCATGATGGAAACCCTGGAACGGCTGCAGCGGAGATACGGCGGAGGACTCGTGC

GGGTGCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGGTGTCCGGCGCCAAGAGCA

ACACCATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGGCAGAATGGACGGCCCCA

GACGGCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGCACCAGAGCCGTGGTGT

CCTGCGCCGAGGCCCCCAACATGAAGATCATCGGCAACCGGATCGAGCGGATCAGAAGCG

AGCACGCCGAGACATGGTTCTTCGACGAGAACCACCCCTACCGGACCTGGGCCTACCACG

GCAGCTACGAGGCCCCTACA

>pr582 C4_pr582 – Blast similarity = 99%

GAGGACTCGTGCGGGTGCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGGTGTCCG

GCGCCAAGAGCAACACCATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGGCAGAA

TGGACGGCCCCAGACGGCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGCACCA

GAGCCGTGGTGTCCTGCGCCGAGGCCCCCAACATGAAGATCATCGGCAACCGGATCGAGC

GGATCAGAAGCGAGCACGCCGAGACATGGTTCTTCGACGAGAACCACCCCTACCGGACCT

GGGCCTACCACGGCAGCTACGAGGCCCCTACACAGGGCAGCGCCAGCAGCCTGATCAACG

GCGTCGTCCGGCTGCTGAGCAAGCCCTGGGATGTGGTCACCGGCGTGACCGGAATCGCCA

TGACCGACACCACCCCCTACGGCCAGCAGCGGGTGTTCAAAGAAAAGGTGGACACCCGGG

TGCCCGACCCCCAGGAAGGCACCAGACAGGTCATGAGCATGGTGTCCAGCTGGCTCTGGA

AAGAGCTGGGCAAGCACAAGCGGCCCAGAGTGTGCACCAAAGAAGAGTTCATCAACAAAG

TGCGGAGCAACGCCGCCCTGGGCGCCATCTTCGAGGAAGAGAAAGAGTGGAAAACCGCCG

TCGAGGCCGTGAACGACCCCCGGTTTTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACC

TGAGAGGCGAGTGCCAGAGCTGCGTGTACAACATGATGGGCAAGCGCGAGAAGAAGCAGG

GCGAGTTCGGCAAGGCCAAGGGCAGCCGGGCCATCTGGTACATGTGGCTGGGCGCCAGAT

TCCTGGAATTTGAGGCCCTGGGCTTCCTGAACGAGGACCACTGGATGGGCAGAGAGAACA

GCGGCGGAGGCGTGGAAGGCCTGGGCCTGCAAAGACTGGGCTACGTGCTGGAAGAGATGA

GCAGAATCCCTGGCGGCAGGATGTACGCCGACGACACC

>pr1260 C4_pr1260 – Blast similarity = 99%

GTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACCTGAGAGGCGAGTGCCAGAGCTGCGT

GTACAACATGATGGGCAAGCGCGAGAAGAAGCAGGGCGAGTTCGGCAAGGCCAAGGGCAG

CCGGGCCATCTGGTACATGTGGCTGGGCGCCAGATTCCTGGAATTTGAGGCCCTGGGCTT

CCTGAACGAGGACCACTGGATGGGCAGAGAGAACAGCGGCGGAGGCGTGGAAGGCCTGGG

CCTGCAGAGACTGGGCTACGTGCTGGAAGAGATGAGCAGAATCCCTGGCGGCAGGATGTA

CGCCGACGACACCGCCGGCTGGGACACCCGGATCAGCAGATTCGACCTGGAAAACGAGGC

CCTGATCACCAACCAGATGGAAAAGGGCCACCGGGCCCTGGCCCTGGCCATCATCAAGTA

CACCTACCAGAACAAGGTGGTCAAAGTGCTGCGGCCTGCCGAGAAGGGCAAGACCGTGGA

TGGACATCATCA

>prMid C4_prMid – Blast similarity = 99%

CTGCCGAGAGGGCAAGACCGTGATGGACATCATCAGCCGGCAGGACCAGCGGGGCAGCGG

CCAGGTGGTCACCTACGCCCTGAACACCTTCACCAACCTGGTGGTGCAGCTGATCCGGAA

CATGGAAGCCGAAGAGGTGCTGGAAATGCAGGACCTGTGGCTGCTGCGGCGGAGCGAGAA

AGTGACCAACTGGCTGCAGTCCAACGGCTGGGACCGGCTGAAGAGAATGGCCGTGTCCGG

CGACGACTGCGTGGTCAAGCCCATCGACGACAGATTCGCCCACGCCCTGCGGTTTCTGAA

CGACATGGGAAAAGTGCGGAAGGACACCCAGGAATGGAAGCCCAGCACCGGCTGGGATAA

CTGGGAGGAAGTGCCCTTCTGCAGCCACCACTTCAACAAGCTGCACCTGAAGGACGGCCG

GTCCATCGTGGTGCCCTGCAGACACCAGGACGAGCTGATCGGCAGAGCCCGGGTGTCACC

TGGCGCCGGATGGTCCATCAGAGAGACAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTG

GCAGCTGCTGTACTTCCACCGGCGGGACCTGCGGCTGATGGCCAACGCCATCTGCAGCAG

CGTGCCCGTGGATTGGGTGCCCACCGGCAGAACCACCTGGTCCATCCACGGCAAGGGCGA

GTGGATGACCACCGAGGACATGCTGGTCGTGTGGAACAGAGTGTGGATCGAAGAGAACGA

CCACATGGAAGATAAGACCCCCGTGACCAAGTGGACCGACATCCCCTACCTGGGCAAGCG

GGAAGATCTGTGGTGCGGCTCCCTGATCGGCCACAGACCTAGGACCACCTGGGCCGAGAA

CATCAAGAACACCGTGAACATGGTCCGACGGATCATCGGCGACGAAAAGAAGTACATGGA

CTACCTGAGCACACAAGT

>pr1980 C4_pr1980 – Blast similarity = 99%

ACGACAGATTCGCCCACGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCGGAAGGACA

CCCAGGAATGGAAGCCCAGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTTCTGCAGCC

ACCACTTCAACAAGCTGCACCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCAGACACC

AGGACGAGCTGATCGGCAGAGCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAGAGAGA

CAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGGCGGG

ACCTGCGGCTGATGGCCAACGCCATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCACCG

GCAGAACCACCTGGTCCATCCACGGCAAGGGCGAGTGGATGACCACCGAGGACATGCTGG

TCGTGTGGAACAGAGTGTGGATCGAAGAGAACGACCACATGGAAGATAAGACCCCCGTGA

CCAAGTGGACCGACATCCCCTACCTGGGCAAGCGGGAAGATCTGTGGTGCGGCTCCCTGA

TCGGCCACAGACCTAGGACCACCTGGGCCGAGAACATCAAGAACACCGTGAACATGGTCC

GACGGATCATCGGCGACGAAGAGAAGTACATGGACTACCTGAGCACACAAGTGCGATACC

TGGGCGAGGAAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAACA

CCGTGCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGGCGAGCAGC

TGGGAGAGACCCTGGGAATGAGCAAAAGCCGCCATCAACAAGCACATCCAGACACTGAGA

GACTGGGGAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCTGAGCCTATC

CAGCTGCTGAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAGCGTGGCCGTGCTG

CCTGTGATCGACTCCACCAATCAGTACCTGCTGGACAGAATCG

>pr2647 C4_pr2647 – Blast similarity = 99%

GCGAGGAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAACACCGT

GCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGGCGAGCAGCTGGG

AGAGACCCTGGGAATGAGCAGAGCCGCCATCAACAAGCACATCCAGACACTGAGAGACTG

GGGAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCTGAGCCTATCCAGCT

GCTGAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAGCGTGGCCGTGCTGCCTGT

GATCGACTCCACCAATCAGTACCTGCTGGACAGAATCGGAGAGCTGAAGTCCGGCGACGC

CTGCATCGCCGAGTACCAGCAGGCTGGCAGAGGAGGCAGAGGACGGAAGTGGTTCAGCCC

ATTCGGAGCCAACCTGTACCTGTCCATGTTCTGGAGACTGGAGCAGGGACCTGCTGCTGC

CATCGGACTGAGTCTGGTGATCGGAATCGTGATGGCCGAGGTGCTGAGAAAGCTGGGAGC

CGACAAGGTGAGAGTGAAGTGGCCTAATGACCTGTACCTCCAGGACCGCAAGCTGGCTGG

CATCCTGGTGGAGCTGACAGGCAAGACAGGCGATGCCGCTCAGATCGTGATCGGAGCCGG

AATCAACATGGCCATGAGAAGAGTGGAGGAGAGCGTGGTGAACCAGGGCTGGATCACCCT

GCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCCGCCATGCTGATCAGAGAGCT

GAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTGAGCAGATGGGA

GAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATCGGCGACAAGGAAATCTT

CGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGGAGCAGGACGGATCA

>prB621 C4_prB62 – Blast similarity = 99%

TGGATCCCCTGCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCCGCCATGCTGA

TCAGAGAGCTGAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTGA

GCAGATGGGAGAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATCGGCGACA

AGGAAATCTTCGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGGAGCAGG

ACGGAATCATCAAGCCCTGGATGGGCGGAGAAATCTCCCTGAGAAGCGAAACAG

C5:

>prStart C5_prStart – Blast similarity = 98%

AGGCGAGACACTGGGCGAGAAATGGAAGGCCCGGCTGAACCAGATGAGCGCCCTGGAGTT

CTACAGCTACAAGAAGTCCGGCATCACCGAAGTGTGCCGGGAAGAGGCCAGACGGGCCCT

GAAGGATGGCGTGGCAACAGGCGGCCACGCCGTGTCTAGAGGCAACGCCAAACTGCGGTG

GC

>pr582 C5_pr582 – Blast similarity = 99%

GGAGACTCGTGCGGGTGCCCCTGAGCAGAAACAGCACCCACGAGATGTACTGGGTGTCCG

GCGCCAAGAGCAACACCATCAAGAGCGTGTCCACCACCAGCCAGCTGCTGCTGGGCAGAA

TGGACGGCCCCAGACGGCCCGTGAAGTACGAAGAGGACGTGAACCTGGGCTCTGGCACCA

GAGCCGTGGTGTCCTGCGCCGAGGCCCCCAACATGAAGATCATCGGCAACCGGATCGAGC

GGATCAGAAGCGAGCACGCCGAGACATGGTTCTTCGACGAGAACCACCCCTACCGGACCT

GGGCCTACCACGGCAGCTACGAGGCCCCTACACAGGGCAGCGCCAGCAGCCTGATCAACG

GCGTCGTCCGGCTGCTGAGCAAGCCCTGGGATGTGGTCACCGGCGTGACCGGAATCGCCA

TGACCGACACCACCCCCTACGGCCAGCAGCGGGTGTTCAAAGAAAAGGTGGACACCCGGG

TGCCCGACCCCCAGGAAGGCACCAGACAGGTCATGAGCATGGTGTCCAGCTGGCTCTGGA

AAGAGCTGGGCAAGCACAAGCGGCCCAGAGTGTGCACCAAAGAAGAGTTCATCAACAAAG

TGCGGAGCAACGCCGCCCTGGGCGCCATCTTCGAGGAAGAGAAAGAGTGGAAAACCGCCG

TCGAGGCCGTGAACGACCCCCGGTTTTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACC

TGAGAGGCGAGTGCCAGAGCTGCGTGTACAACATGATGGGCAAGCGCGAGAAGAAGCAGG

GCGAGTTCGGCAAGGCCAAGGGCAGCCGGGCCATCTGGTACATGTGGCTGGGCGCCAGAT

TCCTGGAATTTGAGGCCCTGGGCTTCCTGAACGAGGACCACTGGATGGGCAGAGAGAACA

GCGGCGGAGGCGTGGAAGGCCTGGGCCTGCAGAGACTGGGCTACGTGCTGGAAGAGATGA

G

>pr1260 C5_pr1260 – Blast similarity = 99%

TTTTGGGCCCTGGTGGACAAAGAGAGAGAGCACCACCTGAGAGGCGAGTGCCAGAGCTGC

GTGTACAACATGATGGGCAAGCGCGAGAAGAAGCAGGGCGAGTTCGGCAAGGCCAAGGGC

AGCCGGGCCATCTGGTACATGTGGCTGGGCGCCAGATTCCTGGAATTTGAGGCCCTGGGC

TTCCTGAACGAGGACCACTGGATGGGCAGAGAGAACAGCGGCGGAGGCGTGGAAGGCCTG

GGCCTGCAGAGACTGGGCTACGTGCTGGAAGAGATGAGCAGAATCCCTGGCGGCAGGATG

TACGCCGACGACACCGCCGGCTGGGACACCCGGATCAGCAGATTCGACCTGGAAAACGAG

GCCCTGATCACCAACCAGATGGAAAAGGGCCACCGGGCCCTGGCCCTGGCCATCATCAAG

TACACCTACCAGAACAAGGTGGTCAAAGTGCTGCGGCCTGCCGAGAAGGGCAAGACCGTG

ATGGACATCATCAGCCGGCAGGACCAGCGGGGCAGCGGCCAGGTGGTCACCTACGCCCTG

AACACCTTCACCAACTGGG

>prMid C5_prMid – Blast similarity = 99%

CAGCAGGATCCGGAACATGGAAGCCGAAGAGGTGCTGGAAATGCAGGACCTGTGGCTGCT

GCGGCGGAGCGAGAAAGTGACCAACTGGCTGCAGTCCAACGGCTGGGACCGGCTGAAGAG

AATGGCCGTGTCCGGCGACGACTGCGTGGTCAAGCCCATCGACGACAGATTCGCCCACGC

CCTGCGGTTTCTGAACGACATGGGAAAAGTGCGGAAGGACACCCAGGAATGGAAGCCCAG

CACCGGCTGGGATAACTGGGAGGAAGTGCCCTTCTGCAGCCACCACTTCAACAAGCTGCA

CCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCAGACACCAGGACGAGCTGATCGGCAG

AGCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAGAGAGACAGCCTGCCTGGCCAAGAG

CTACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGGCGGGACCTGCGGCTGATGGCCAA

CGCCATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCACCGGCAGAACCACCTGGTCCAT

CCACGGCAAGGGCGAGTGGATGACCACCGAGGACATGCTGGTCGTGTGGAACAGAGTGTG

GATCGAAGAGAACGACCACATGGAAGATAAGACCCCCGTGACCAAGTGGACCGACATCCC

CTACCTGGGCAAGCGGGAAGATCTGTGGTGCGGCTCCCTGATCGGCCACAGACCTAGGAC

CACCTGGGCCGAGAACATCAAGAACACCGTGAACATGGTCCGACGGATCATCGGCGACGA

AGAGAAGTACATGGACTACCTGAGCACACAAGTGCGATACCTGGGGCGAGGAAGGCAGCA

CCCCCCGGAGT

>pr1980 C5_pr1980 – Blast similarity = 99%

CGACGACAGATTCGCCCACGCCCTGCGGTTTCTGAACGACATGGGAAAAGTGCGGAAGGA

CACCCAGGAATGGAAGCCCAGCACCGGCTGGGATAACTGGGAGGAAGTGCCCTTCTGCAG

CCACCACTTCAACAAGCTGCACCTGAAGGACGGCCGGTCCATCGTGGTGCCCTGCAGACA

CCAGGACGAGCTGATCGGCAGAGCCCGGGTGTCACCTGGCGCCGGATGGTCCATCAGAGA

GACAGCCTGCCTGGCCAAGAGCTACGCCCAGATGTGGCAGCTGCTGTACTTCCACCGGCG

GGACCTGCGGCTGATGGCCAACGCCATCTGCAGCAGCGTGCCCGTGGATTGGGTGCCCAC

CGGCAGAACCACCTGGTCCATCCACGGCAAGGGCGAGTGGATGACCACCGAGGACATGCT

GGTCGTGTGGAACAGAGTGTGGATCGAAGAGAACGACCACATGGAAGATAAGACCCCCGT

GACCAAGTGGACCGACATCCCCTACCTGGGCAAGCGGGAAGATCTGTGGTGCGGCTCCCT

GATCGGCCACAGACCTAGGACCACCTGGGCCGAGAACATCAAGAACACCGTGAACATGGT

CCGACGGATCATCGGCGACGAAGAGAAGTACATGGACTACCTGAGCACACAAGTGCGATA

CCTGGGCGAGGAAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAA

CACCGTGCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGGCGAGCA

GCTGGGAGAGACCCTGGGAATGACAAAAAGCCGCCATCAACAAGCACATCCAGACACTGA

GAGACTGGGGAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCTGAGCCTA

TCCAGCTGCTGAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAGCGTGGCCGTGC

TGCCTGTGATCGACTCCACCAATCAGTACCTGCTGGACAGAATCGG

>pr2647 C5_pr2647 – Blast similarity = 99%

TGGGCGAGGAGGCAGCACCCCCGGAGTGCTGGCAGGAAAGCTTGCAGGAAAGGACAACAC

CGTGCCCCTGAAGCTGATCGCCCTGCTGGCCAACGGCGAGTTCCACTCTGGCGAGCAGCT

GGGAGAGACCCTGGGAATGAGCAGAGCCGCCATCAACAAGCACATCCAGACACTGAGAGA

CTGGGGAGTGGACGTGTTCACCGTGCCTGGCAAGGGCTACAGCCTGCCTGAGCCTATCCA

GCTGCTGAACGCCAAGCAGATCCTGGGACAGCTGGATGGCGGAAGCGTGGCCGTGCTGCC

TGTGATCGACTCCACCAATCAGTACCTGCTGGACAGAATCGGAGAGCTGAAGTCCGGCGA

CGCCTGCATCGCCGAGTACCAGCAGGCTGGCAGAGGAGGCAGAGGACGGAAGTGGTTCAG

CCCATTCGGAGCCAACCTGTACCTGTCCATGTTCTGGAGACTGGAGCAGGGACCTGCTGC

TGCCATCGGACTGAGTCTGGTGATCGGAATCGTGATGGCCGAGGTGCTGAGAAAGCTGGG

AGCCGACAAGGTGAGAGTGAAGTGGCCTAATGACCTGTACCTCCAGGACCGCAAGCTGGC

TGGCATCCTGGTGGAGCTGACAGGCAAGACAGGCGATGCCGCTCAGATCGTGATCGGAGC

CGGAATCAACATGGCCATGAGAAGAGTGGAGGAGAGCGTGGTGAACCAGGGCTGGATCAC

CCTGCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCCGCCATGCTGATCAGAGA

GCTGAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTGAGCAGATG

GGAGAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATCGGCGACAAGGAAAT

CTTCGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGGAGCAGGACGGAAT

CATCAAGCCCTGGATGGCGGA

>prB621 C5_prB62 – Blast similarity = 99%

CTGGATCCCCTGCAGGAGGCTGGCATCAACCTGGACCGGAACACCCTGGCCGCCATGCTG

ATCAGAGAGCTGAGAGCCGCTCTGGAGCTGTTCGAGCAGGAGGGACTGGCTCCTTACCTG

AGCAGATGGGAGAAGCTGGACAACTTCATCAACAGACCTGTGAAGCTGATCATCGGCGAC

AAGGAAATCTTCGGCATCTCCAGAGGAATCGACAAGCAGGGAGCTCTGCTGCTGGAGCAG

GACGGAATCATCAAGCCCTGGATGGGCGGAGAAATCTCCCTGAGAAGCGCGAAGA

6.3: Appendix C – CD